Bashiri and Moslemi *Journal of Industrial Engineering International* 2013, **9**:2
http://www.jiei-tsb.com/content/9/1/2

Journal of Industrial Engineering International
a SpringerOpen Journal

ORIGINAL RESEARCH Open Access

The analysis of residuals variation and outliers to obtain robust response surface

Mahdi Bashiri* and Amir Moslemi

Abstract

In this paper, the main idea is to compute the robust regression model, derived by experimentation, in order to achieve a model with minimum effects of outliers and fixed variation among different experimental runs. Both outliers and nonequality of residual variation can affect the response surface parameter estimation. The common way to estimate the regression model coefficients is the ordinary least squares method. The weakness of this method is its sensitivity to outliers and specific residual behavior, so we pursue the modified robust method to solve this problem. Many papers have proposed different robust methods to decrease the effect of outliers, but trends in residual behaviors pose another important issue that should be taken into account. The trends in residuals can cause faulty estimations and thus faulty future decisions and outcomes, so in this paper, an iterative weighting method is used to modify both the outliers and the residuals that follow abnormal trends in variation, like descending or ascending trends, so they will have less effect on the coefficient estimation. Finally, a numerical example illustrates the proposed approach.

Keywords: Coefficient estimation, Response surface, Ordinary least squares, Outliers, Robust model

Introduction

In many cases, especially in experimental results, some of the data are wrong and should be treated as outliers. These points, which may occur because of operator reading faults and the like, may have a confusing effect on the total interpretation of the results. A common method of explaining and analyzing the results of experiments is by response surface design. This term is used for a regression equation that shows the whole behavior of the control variables, the nuisance factors, and the response or responses. We can use the estimated function to predict the response to changes in the values of specific controllable factors. After determining an experimental design and performing experiments, the next steps are generally statistical analysis and then the selection of values for the input variables so as to optimize the output. This can be done by fitting a regression model between the controllable factors and the response variables. Future interpretations are based on this regression model, so the exact model is very important and may affect the optimization stage. This model is generally constructed by the ordinary least squares (OLS) method. But basic OLS is very sensitive to outliers, and they may have an inordinate effect on the ultimate conclusion. So a robust method or a modified OLS should be used for decreasing the outliers' sensitivity.

Our goal in this study is to decrease the destructive effect of outliers. In order to do so, at the first stage, the robust regression or modified regression model should be computed. Then the trend of residuals in response surface design is another aspect which should be considered. The trend behaviors among residuals, both descending and ascending trends, can cause faulty interpretations. However, there are some assumptions in estimating the regression coefficients that should not be violated. It seems that by decreasing the effects of some residuals in the coefficient estimation stage, the initial assumptions can be satisfied, and moreover, because this decreases the overall variability, the robustness of the model will be increased. The main purpose of this paper is to decrease the effects of outliers that violate the variance equality test of residuals. An example of such trends and abnormal behavior is shown in Figure 1.

As mentioned before, the OLS method is very sensitive to outliers. To diminish the effect of these points, some alternative methods of model fitting, such as least absolute

* Correspondence: bashiri@shahed.ac.ir
Department of Industrial Engineering, Faculty of Engineering, Shahed University, Tehran, Iran

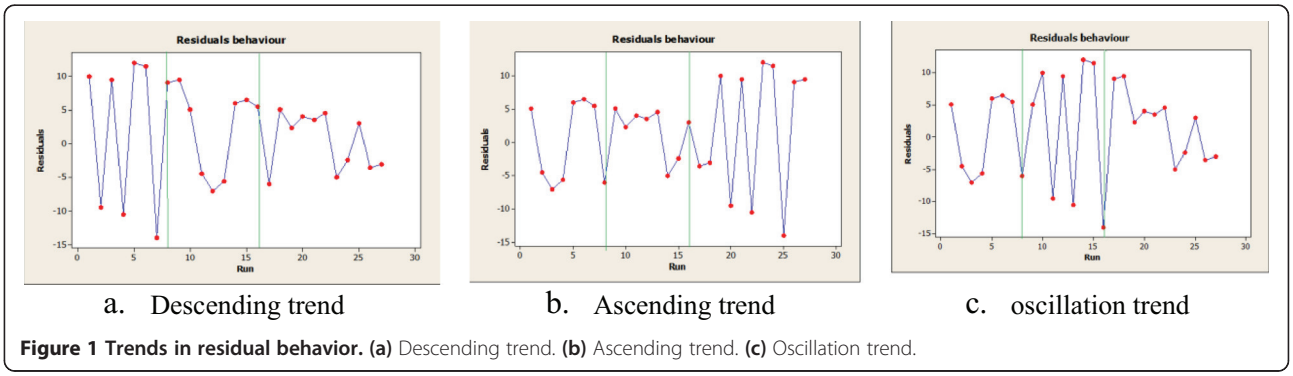

Figure 1 Trends in residual behavior. (a) Descending trend. **(b)** Ascending trend. **(c)** Oscillation trend.

deviations and other robust approaches that simplify the task of outlier identification by weighting the large residuals, are used instead of OLS. Response surfaces have been studied by many researchers, and many approaches have been proposed either to obtain efficient response surfaces or optimize the response surface by different models. Hejazi et al. (2010) proposed a novel approach based on goal programming to find the best combination of factors to optimize multi-response-multi-covariate surfaces by considering location and dispersion effects. Kazemzadeh et al. (2008) proposed a method to optimize multi-response surfaces based on a goal programming method. Robust regression approaches have also been surveyed by many researchers. Huber (Bertsimas and Shioda 2007) proposed M-estimator methods to obtain robust regression. Morgenthaler and Schumacher (1999) discussed robust response surfaces in chemistry based on design of experiment. Because of the weakness of the previous methods in compensating for outliers, the redescending M-estimators (also named GM-estimators) were proposed by Andrews et al. (1972), which are able to reject extreme outliers entirely. Hund et al. (2002) presented various methods of outlier detection and evaluated robustness tests with different experimental designs. Bickela and Frühwirthb (2006) compared different robust estimators with their applications. The M- and GM-estimators work by an iterative procedure. As a consequence, several authors (e.g., Cummins and Andrews 1995) have called these estimators as iteratively reweighted least squares, or IRLS methods. Ortiz et al. (2006) discussed some of the robust methods used for robust regression in analytical chemistry. To obtain a more efficient and yet robust method, Siegel (1982) proposed the repeated median estimator. Also another useful robust method is the least median squared (LMS) method proposed by Rousseeuw (1984). Massart et al. (1986) showed the advantages of its use in chemical analysis. The other useful method is the least trimmed squared (LTS) that was proposed by Rousseeuw and Leroy (1987). Nguyena and Welsch (2010) studied outlier detection and proposed a new least trimmed squares approximation. Both the LMS and the LTS are defined by minimizing a robust measure of the scatter of the residuals. Generalizing this, Rousseeuw and Yohai (1984) introduced S-estimators which are significantly more efficient than the previous estimators. A more recent suggestion is the constrained M-estimates, or CM, proposed by Mendes and Tyler (1995), which combines the good local properties of the M-estimates and the good global robustness properties of the S-estimates. A 'partial' version of the M-estimator based on the 'fair' ψ function and an appropriate weighting scheme was recently proposed by Serneels et al. (2005). The authors believed that the partial robust M-regression outperforms existing methods for robust partial least square regression. Bertsimas and Shioda (2007) presented mixed integer programming or MIP models for the classification and robust regression problems. Zioutas and Avramidis (2005) examined the effect of deleting outliers in the regression model obtained by mixed integer programming and compared the performance of this model with that of least squares, or LS, and LMS. Another new method in robust regression is the mixed linear model surveyed by Dornheim and Brazauskas (2011). (Pop and Sârbu 1996) proposed a new fuzzy regression algorithm to obtain robust models. Maronna et al. (2006) proposed many M-estimators using robust regression methods in both single response and multiple responses. Shahriari et al. (2011) proposed a novel two-step robust estimation of the process mean method based on M-estimator and their method is less sensitive to the presence of outliers. For better illustration of proposed method, the literature review has been classified in Table 1.

In this paper, a novel robust approach considering both outlier data and trends in residuals variations which do not violate the normality assumption is discussed.

This paper is organized as follows. The section 'Robust estimation of the coefficients by iterative weighting methods' presents the robust modification of the response surface by an iterative weighting procedure. The proposed method is defined in section 'Robust estimation of coefficients by testing equality of variations in specified intervals'. To illustrate the proposed method, a numerical

Table 1 Summary of literature review

Author (year)	Characteristic				
	Iterative method	Considering equality of variation of residuals	Ordinary least square	Different M-estimators	Other regression models
Huber (1981)	✓			✓	
Siegel (1982)	✓			✓	
Rousseeuw (1984)	✓			✓	
Massart et al. (1986)	✓			✓	
Rousseeuw and Leroy (1987)	✓			✓	
Cummins and Andrews (1995)	✓		✓		
Pop and Sârbu (1996)					✓
Morgenthaler and Schumacher (1999)	✓		✓	✓	
Hund et al. (2002)			✓		
Zioutas and Avramidis (2005)					✓
Serneels et al. (2005)	✓			✓	
Bickela and Frühwirthb (2006)	✓			✓	
Ortiz et al. (2006)	✓		✓	✓	
Bertsimas and Shioda (2007)					✓
Nguyena and Welsch (2010)	✓			✓	
Dornheim and Brazauskas (2011)					✓
Proposed REVIM	✓	✓	✓		

example is presented in section 'Numerical example'. Finally, the last section is the 'Conclusion' of this paper.

Robust estimation of the coefficients by iterative weighting methods

To compensate for the effects of the outlier values, we can either remove the outlier data or modifying them by weighting the residuals. The first approach is not rational, so we choose to modify them in order to decrease the effect of outliers in the coefficient estimation stage. The proposed idea is as follows:

$$E(y_i) = \mu_i\left(\beta_1, \ldots \beta_p\right). \tag{1}$$

In this equation, μ_i is a function defined by unknown coefficients (β_i). For example, if $\mu_1 = \beta_1 + \beta_2 x_1$ and x_i are constants, the response y_i can be obtained from the experimental results, and the regression model describes the relation between the variables and the expected values of the y_i.

If all the measurements are good, then the OLS method provides a reasonable model and the coefficients are estimated by minimizing the following equation:

$$\begin{aligned}&\left(y_1 - \mu_1\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)\right)^2 + \ldots \\ &+\left(y_n - \mu_n\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)\right)^2.\end{aligned} \tag{2}$$

However, if the results appear abnormal, which may be a consequence of residual behavior in the experiments, the coefficients are determined by minimizing the following equation. The abnormality occurs when a residual behaves like an outlier:

$$\begin{aligned}&w_1\left(y_1 - \mu_1\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)\right)^2 + \ldots \\ &+w_n\left(y_n - \mu_n\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)\right)^2\end{aligned} \tag{3}$$

The weights are not pre-assigned values because the quality of each y_i is not known in advance. The reasonable values for the weights are based on the residuals defined by the following equation:

$$r_i = y_i - \mu_i\left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right). \tag{4}$$

To make the estimator invariant with respect to the scale of the residuals, the r_i is divided by 's,' which is a robust estimation of the scale. The value of 's' is often taken to be equal to 1.4826 MAD, where MAD is the median of the absolute deviations of the residuals from their median and 1.4826 is a bias adjustment for the standard deviation under the normal distribution.

The weights should be inversely proportional to the value of the residuals, $w_i = \frac{c}{|r_i|}$. In other words, the residuals with large values are weighted less, and this method produces better coefficient estimates. These

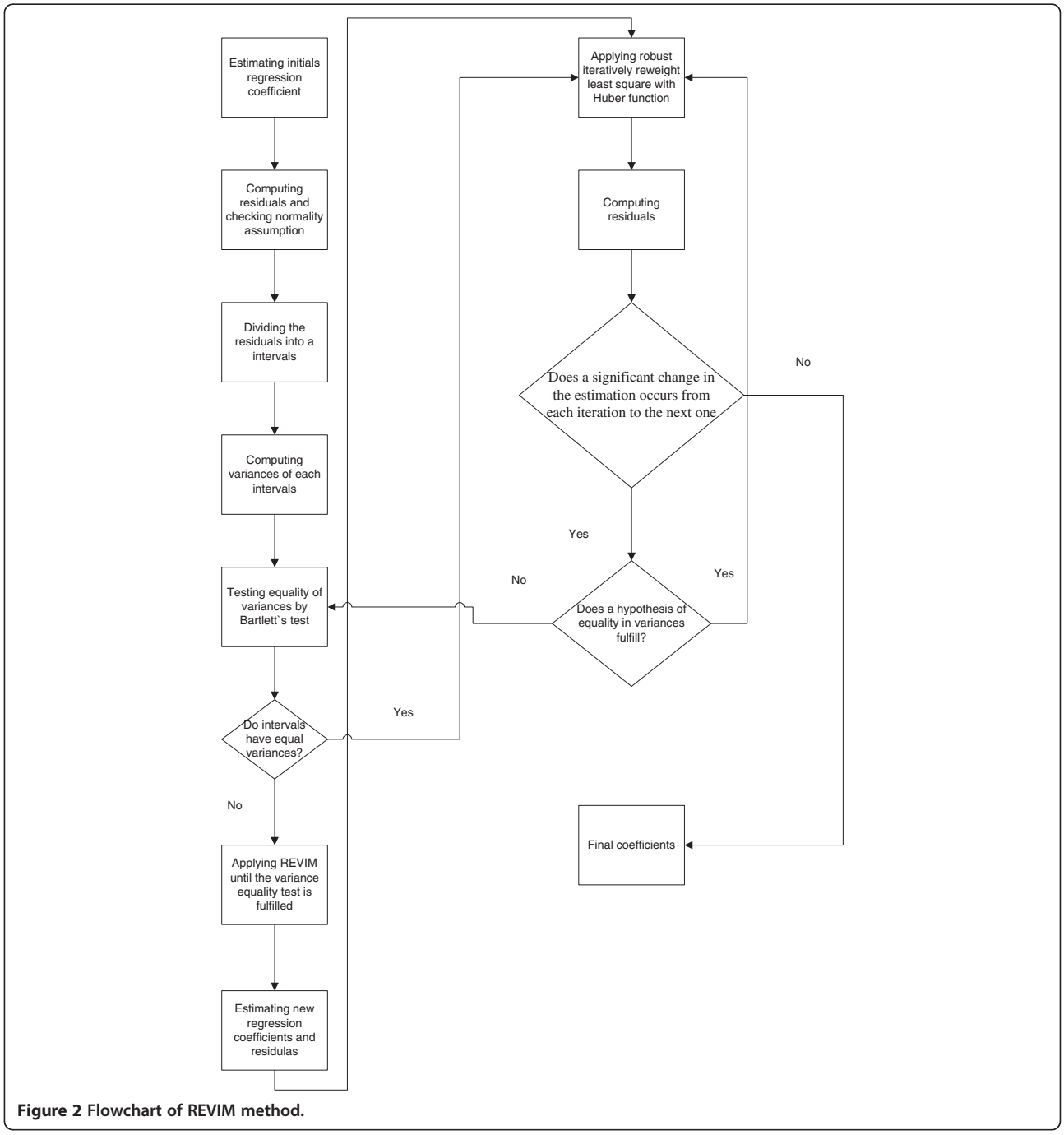

Figure 2 Flowchart of REVIM method.

weights can be chosen by a function such as the Huber weight function:

$$\begin{cases} w_i = 1 & if\,|r_i| < c \\ w_i = \dfrac{c}{|r_i|} & f|r_i| > c \end{cases}, \tag{5}$$

where c is a constant. The procedure is as follows: compute the first coefficients of the regression model, compute the residuals and weights, and then compute the new coefficients by the equation. This procedure can be repeated until a good solution is obtained, because the values of the coefficients and the values of the residuals and weights are different. This procedure is known as iterative weighting OLS. The procedure terminates when the change in the estimation from one iteration to the next is sufficiently small.

This iterative method is good for modifying outliers, but the trends in residual behavior are not considered. As illustrated before, another approach, in addition to taking outliers into account, is the equality of variation

Table 2 A hypothetical data created according to Box-Behnken design

Material 1	Material 2	Material 3	Material 4	Block	*Y* value	x_1	x_2	x_3	x_4
0.82	1	−55	0	1	96.49	−0.1	−0.1	0	0
0.82	1	−45	−25	1	93.22	−0.1	−0.1	1	−1
0.82	1	−65	−25	1	87.43	−0.1	−0.1	−1	−1
0.91	0.91	−55	0	1	77.20	0.8	−1	0	0
0.91	1.09	−55	0	1	82.83	0.8	0.8	0	0
0.73	0.91	−55	0	1	94.87	−1	−1	0	0
0.73	1.09	−55	0	1	63.46	−1	0.8	0	0
0.82	1	−65	25	1	91.88	−0.1	−0.1	−1	1
0.82	1	−45	25	1	91.88	−0.1	−0.1	1	1
0.83	1.02	−55	0	2	100.28	0	0.1	0	0
0.92	1.02	−55	−25	2	90.44	0.9	0.1	0	−1
0.74	1.02	−55	−25	2	92.53	−0.9	0.1	0	−1
0.83	0.93	−65	0	2	90.32	0	−0.8	−1	0
0.83	1.11	−65	0	2	91.45	0	1	−1	0
0.83	1.11	−45	0	2	90.85	0	1	1	0
0.83	0.93	−45	0	2	69.08	0	−0.8	1	0
0.92	1.02	−55	25	2	88.55	0.9	0.1	0	1
0.74	1.02	−55	25	2	91.55	−0.9	0.1	0	1
0.83	1.02	−55	0	3	90.96	0	0.1	0	0
0.83	0.92	−55	−25	3	85.67	0	−0.9	0	−1
0.83	1.11	−55	−25	3	80.21	0	1	0	−1
0.74	1.02	−45	0	3	84.74	−0.9	0.1	1	0
0.93	1.02	−65	0	3	93.70	1	0.1	−1	0
0.74	1.02	−65	0	3	92.24	−0.9	0.1	−1	0
0.93	1.02	−45	0	3	95.60	1	0.1	1	0
0.83	1.11	−55	25	3	87.83	0	1	0	1
0.83	0.92	−55	25	3	83.21	0	−0.9	0	1

between residuals which is the main idea of the rest of this paper.

Robust estimation of coefficients by testing equality of variations in specified intervals

First, normality assumption is checked. If the normality assumption is violated, this robust approach based on Huber function cannot be applied. The method proposed in this part begins by dividing the experiment runs into *a* intervals to examine the hypothesis of equality in variations in these intervals. The equality test used in this paper is Bartlett's test (Anderson and McLean 1974). The number of points in each interval can be chosen in the analysis stage. This stage satisfies one of the OLS hypotheses. If this parameter is small, the variation between points might be large and if the number of the points is large, the equality test of the variances may not be reliable, so this value should be determined rationally.

Bartlett's test

Although residual plots are frequently used to diagnose inequality of variance, statistical tests have also been proposed. One widely used procedure is Bartlett's test. The procedure involves computing a statistic whose sampling distribution is closely approximated by the chi-square distribution with *a-1* degrees of freedom, where the *a* random samples are from independent normal populations. This statistic is defined as

$$x^2 = 2.3026\frac{q}{c} \tag{6}$$

where

$$q = (N-a)\log_{10}^{s_p^2} - \sum_{i=1}^{a}(n_i-1)\log_{10}^{s_i^2}, \tag{7}$$

$$c = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}(n_i-1)^{-1} - (N-a)^{-1}\right), \tag{8}$$

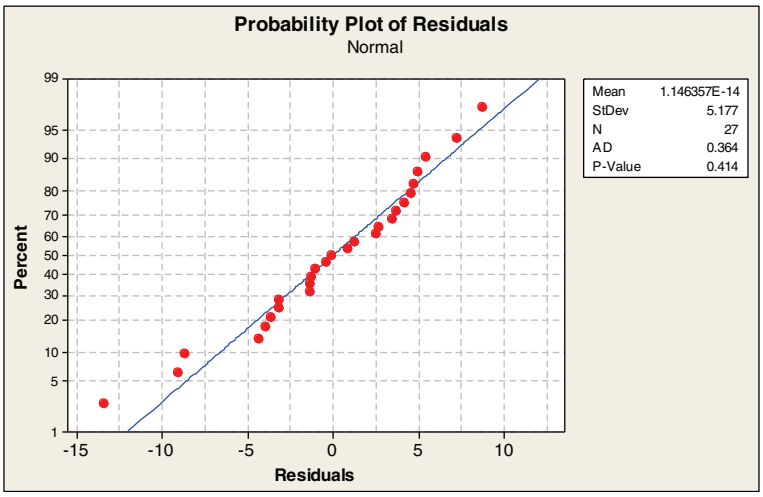

Figure 3 Normality assumption and adequacy checking of residuals.

$$s_p^2 = \frac{\sum_{i=1}^{a}(n_i - 1)s_i^2}{N - a}, \tag{9}$$

and s_i^2 is the sample variance of the ith population.

The hypothesis of equality of variances is rejected if $x_0^2 > x_{\alpha,a-1}^2$, where $x_{\alpha,a-1}^2$ is the upper α percentiles of the chi-square distribution with a-1 degrees of freedom.

Table 3 The ANOVA results

Term	Coefficient	Standard error	*T* value	*P* value
Constant	93.6635	4.732	19.794	0
Block 1	−1.6534	2.305	−0.717	0.49
Block 2	1.1817	2.284	0.517	0.616
x_1	0.329	2.636	0.125	0.903
x_2	0.408	2.636	0.155	0.88
x_3	−1.9282	2.416	−0.798	0.443
x_4	0.3978	2.416	0.165	0.872
$x_1 * x_1$	−3.898	4.215	−0.925	0.377
$x_2 * x_2$	−9.5556	4.173	−2.29	0.045
$x_3 * x_3$	−0.1949	3.571	−0.055	0.958
$x_4 * x_4$	−0.9236	3.578	−0.258	0.802
$x_1 * x_2$	12.3039	5.054	2.434	0.035
$x_1 * x_3$	2.0222	4.366	0.463	0.653
$x_1 * x_4$	−0.3262	4.619	−0.071	0.945
$x_2 * x_3$	4.733	4.564	1.037	0.324
$x_2 * x_4$	2.4782	4.342	0.571	0.581
$x_3 * x_4$	−1.4475	4.174	−0.347	0.736

Proposed robust approach (robust equal variances iterative method)

The following steps are proposed as an iterative method to decrease the effect of trends in the residuals and improve the robust estimation of coefficients. The proposed model is based on OLS method which should be modified.

First of all, our goal is that the residuals that violate the hypothesis of equality in variances should have less effect on the estimates of the coefficients of the regression model, so we should consider modifying these points to have equal variances. Therefore, the residuals derived by experiments are divided into a intervals, and then the variances of each interval are calculated and denoted by s_i^2. The next step is to test the equality of variances with Bartlett's test; if the result shows that the variances in a intervals do not have significant differences, this part of the procedure is stopped, whereas if the result shows that the

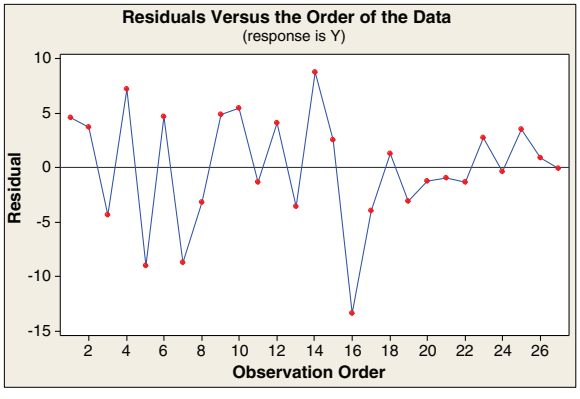

Figure 4 The least squares residuals of the model in order of the runs.

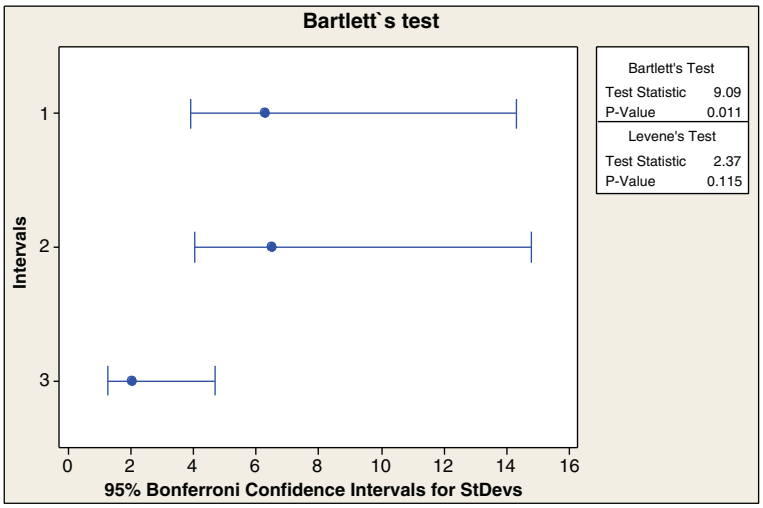

Figure 5 The results of equality test of variances.

variances in the a intervals differ significantly, the iterative weighting procedure is used to modify. As the next step, the critical q statistic in Bartlett's test, for which the hypothesis will be rejected, is computed. The critical q for Bartlett's test is denoted by q_c, and in our case, $q > q_c$. The $s^2_{(1)}$ is the ith rank-sorted variances of the intervals, in descending order. The $s^{2(j)}_{(i)}$ is the variances of the points in ith ranked interval in the jth. In the proposed approach, in higher and higher iterations approach, the residual variances approach equality. The maximum feasible variance of each intervals for which Bartlett's test is fulfilled is denoted as s^2_i max.

Because the q value is greater than q_c after its computation, it should be decreased iteratively until the both values are equal. To do this, we select the largest variance of all intervals. This value is decreased in each iteration. If the value of $s^2_{(1)}$ equals $s^2_{(2)}$ in this decreasing procedure, both $s^2_{(1)}$ and $s^2_{(2)}$ decrease in the next iteration. If the value q does not equal q_c, the values of the variances decrease again. If the values of $s^2_{(1)}$ and $s^2_{(2)}$ equal $s^2_{(3)}$ in this decreasing procedure, all three variances decrease in the next iteration, and this procedure continues until q is equal to q_c. All maximum feasible variances of intervals that satisfy the hypothesis of Bartlett's test are computed with this iterative procedure. Next, we consider two parallel lines l_1:$y = +\omega$ and l_2:$y = -\omega$, with slope zero and parallel to the x-axis. We decrease the ω values iteratively and compute the variances of points between these two lines. Until the variances of the points are equal to the maximum variance of intervals derived from the last step, this decreasing procedure continues. After that, the points outside these lines are weighted by function in (10).

The pseudo code of the proposed method illustrates this approach:

1. Divide residuals into a intervals.
2. Compute the variances of outliers in each interval, and denote these variances by s^2_i.
3. Sort the variances of each interval in descending order and denote them by $s^2_{(i)}$.
4. Calculate the critical value of Bartlett's test in terms of confidence level and the number of intervals.
5. Compute the q value by formula in (7).

Table 4 Bartlett's test results (test statistic = 9.09)

Intervals	Number	Lower	Standard deviation	Upper
1	9	3.94377	6.32629	14.3264
2	9	4.07722	6.54037	14.8112
3	9	1.29441	2.07639	4.7021

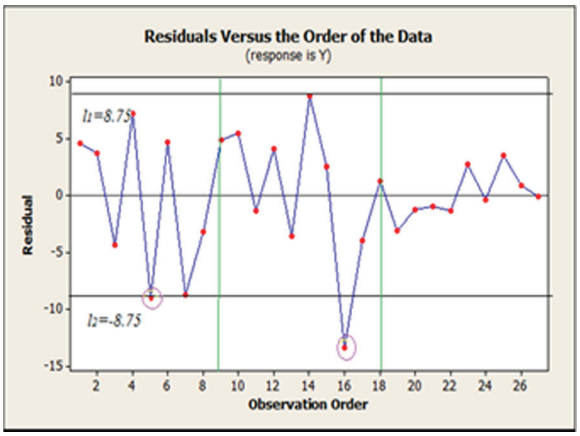

Figure 6 Proposed method.

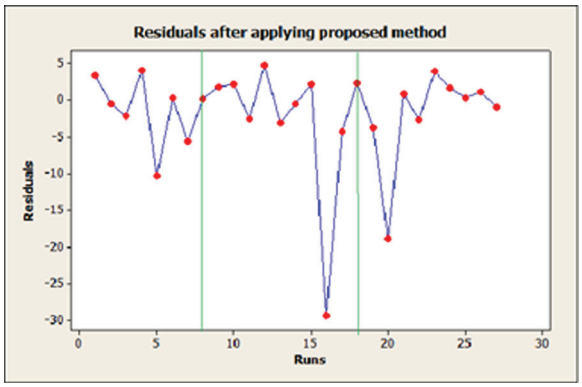

Figure 7 Proposed method residuals.

6. Compute the statistic of Bartlett's test by formula in (6).
7. Compute the q_c by considering the critical value of Bartlett's test based on formula (6).
8. Do while $q \geq q_c$,

$$s_{(1)}^{2(j+1)} = s_{(1)}^{2(j)} - \delta \quad If \quad s_{(i)}^{2(j)} > s_{(2)}^{2(j)}$$
$$s_{(1)}^{2(j+1)} = s_{(1)}^{2(j+1)} = \ldots = s_{(f)}^{2(j+1)} = s_{(1)}^{2(j)} - \delta \quad If \quad s_{(1)}^{2(j)} \leq s_{(f)}^{2(j)} \; f = 2, 3, \ldots$$
$$j = j + 1.$$

9. Determine s_i^2 max as the maximum feasible variances of each intervals to satisfy the hypothesis of equal variances.
10. Consider two parallel lines, l_1, l_2, $l_1: y = +\omega$ and $l_2: y = -\omega$.
11. Do while $s_k^2 < s_i^2$ max, $l_1: y = +m\omega$ and $l_2: y = -m\omega$.
12. Determine the points outside lines $l_1$and l_2.

Table 5 Final robust coefficients

Coefficient	Value
Intercept	94.19
x_1	0.63
x_2	1.6
x_3	−1.5
x_4	0.39
x_1x_2	13.29
x_1x_3	2.06
x_1x_4	−0.32
x_2x_3	3.59
x_2x_4	2.46
x_3x_4	−1.44
x_1^2	−4.13
x_2^2	−9.32
x_3^2	−0.19
x_4^2	−1.47
Block 1	−1.36
Block 2	1.28

13. Determine the weight of the outside residuals by function below (formula 10).

$$\begin{cases} w_i = 1 & if\, l_2 < r_i < l_i \\ w_i = \dfrac{1}{|r_i|} & f\, r_i > l_1, r_i < l_2 \end{cases} \tag{10}$$

After this step, by the robust iterative weighting method, the outliers are modified. After each iteration, the test of equality in variances is checked, and if the

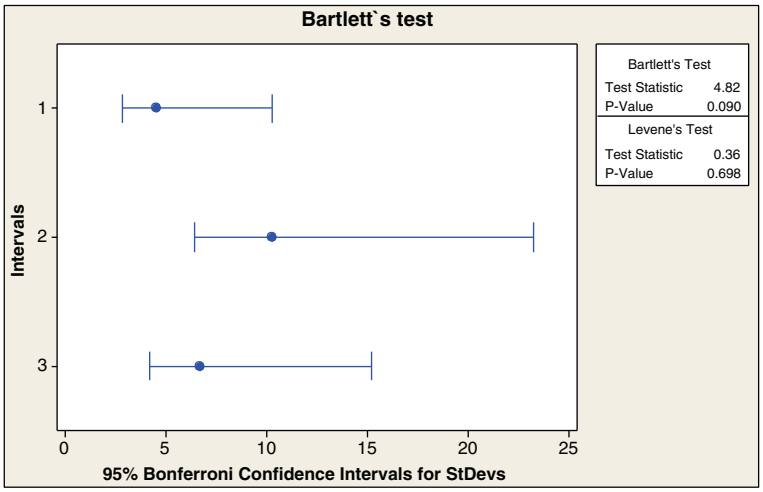

Figure 8 Bartlett's test results after applying REVIM.

Table 6 Final residuals of robust model

Run	Residual
1	3.89
2	3.09
3	−3.94
4	8.07
5	−11.6
6	4.44
7	−9.28
8	−2.75
9	4.28
10	4.73
11	−1.69
12	4.51
13	−2.1
14	5.97
15	1.2
16	−14.5
17	−4.28
18	1.65
19	−3.38
20	−13.6
21	−2
22	−1.29
23	2.63
24	0.19
25	2.71
26	−0.13
27	1.23

hypothesis is violated, the above mentioned method is applied. After the equality of variances is satisfied, the iterative weighting method continues, and this procedure continues as long as a minor change in the estimation occurs from each iteration to the next one. This procedure is called the robust equal variances iterative method procedure (REVIM).

The flowchart in Figure 2 illustrates this method.

Numerical example

This is a hypothetical numerical experiment. Suppose that we have an experiment containing one response variable and four explanatory control variables, each of which has three levels, and the objective of the study is to optimize the yield of a product. The data to be used are shown in Table 2. We want to explore the yield response surface by using a second-order regression model. A Box-Behnken design with 27 treatments is used for this experiment. The blocking is used to diminish the effect of nuisance factors, and the blocks are assigned, for example, to 3 days.

Table 7 Comparison between REVIM approach and OLS method

Coefficient	Method		
	Actual	OLS	REVIM
Intercept	95.67	93.66	94.19
x_1	1.25	0.32	0.63
x_2	−1.16	0.40	1.6
x_3	0.79	−1.92	−1.5
x_4	0.39	0.39	0.39
x_1x_2	14.73	12.30	13.29
x_1x_3	2.29	2.02	2.06
x_1x_4	−0.32	−0.32	−0.32
x_2x_3	−2.66	4.73	3.59
x_2x_4	2.47	2.47	2.46
x_3x_4	−1.44	−1.44	−1.44
x_1^2	−5.4	−3.89	−4.13
x_2^2	−7.71	−9.55	−9.32
x_3^2	1.11	−0.19	−0.19
x_4^2	−3.48	−0.92	−1.47
Block 1	−1.99	−1.65	−1.36
Block 2	3.03	1.18	1.28

The primary fitted response regression model is as follows:

$$\begin{aligned}\hat{y} = {} & 93.65 + 0.32x_1 + 0.4x_2 - 1.92x_3 + 0.39x_4 \\ & - 3.89x_1^2 - 9.55x_2^2 - 0.19x_3^2 - 0.92x_4^2 + 12.3x_1x_2 \\ & + 2.02x_1x_3 - 0.32x_1x_4 + 4.73x_2x_3 + 2.47x_2x_4 \\ & - 1.44x_3x_4 + \text{block effect}.\end{aligned} \tag{11}$$

The normality assumptions are checked in this example and the results are given in Figure 3. The P value obtained by normality test is 0.414. This value shows that the residuals follow normal distribution.

The analysis of variance (ANOVA) results are shown in Table 3.

Figure 4 shows the residuals of the model in the order of runs. As shown in the figure, the residuals have a rough trend, and if we divide the runs into 3 equal intervals, each containing 9 runs, the second interval has larger variance. This can be proved by Bartlett's test.

The result of Bartlett's test is illustrated in Figure 5 and Table 4.

In this case, we have three intervals, so a-1 = 2. If we consider the significance level 0.95, the critical statistic $x^2_{0.05,2}$ is equal to 5.99 and the test statistic = 9.09 is greater than 5.99, then the hypothesis is rejected. Therefore, we want to compute the maximum standard deviation of each interval that satisfies the hypothesis of Bartlett's test. By the proposed method, the maximum values of these standard variations are ordered as follows: $s_1 = 2.07$, $s_2 = 4.95$, $s_3 = 4.95$. Based on these values, we can compute the limits, $l_1 = 8.7$ and $l_1 = -8.75$. Two

points 5 and 16 are outside these limits (circled in the figure), so the weighting procedure is applied and new coefficients are calculated by the robust weighting method. Figure 6 shows the method graphically.

The residuals are computed by this procedure, and the hypothesis of equal variance in three intervals is satisfied. Bartlett's test is applied to the residuals obtained from the proposed method, and the results are given in Figure 7.

The value of Bartlett's test is 4.32, and the hypothesis of equal variances is not rejected. Figure 8 illustrates the result better.

Therefore, the iterative weighting method based on these coefficients to modify the effect of outliers by Huber function with $c = 2$ is applied. This process is applied in each iteration, and if the equality test is not satisfied, the modification is applied. The final robust coefficients and residuals are presented in Tables 5 and 6.

So the residuals in final iteration show that residuals which hinted to be an outlier in first iteration are really outliers, and the model estimation is more accurate and close to the model with no outlier data that we call it *actual model*. These results can be compared with the results obtained by the model with no outlier data. The comparison shows that the proposed model is more precise and accurate than the OLS method in estimation of the regression coefficients by considering unequal variation between residuals. The results are given in Table 7.

Conclusions

A robust estimation of the response surface is the primary goal of this paper. To this end, the proposed method is defined, instead of the common ordinary least squares method of estimating coefficients of the response surface, to decrease the effects of two main causes of the imprecise estimation of coefficients, outliers, and trends in residuals. As the effect of trends in residuals should be taken into account, the proposed method simultaneously modifies the effects of trends and outliers. For each iteration, an equality test of residual variances is performed, and after this hypothesis is satisfied, the outliers are modified. A goal for future research may be to examine the weighting; instead of computing the distance of the residuals from base line after plotting the responses in a normal probability plot (NPP), the weighting function may be proportional to the distance of the response from the NPP regression line.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MB has worked on the literature and modeling, he also proposed response surface methodology. AM has formulated the robust concept and numerical example. Both authors read and approved the final manuscript.

Authors' information

MB is an associate professor of industrial engineering and his research interests are multiple response optimization and facilities location problem. AM is a Ph.D. student of industrial engineering at Amirkabir University of Technology (Tehran Polytechnic).

Acknowledgments

Authors are thankful to the reviewers for their valuable comments.

Received: 1 July 2011 Accepted: 5 March 2012
Published: 13 March 2013

References

Anderson VL, McLean RA (1974) Design of experiments: a realistic approach. Dekker, New York
Andrews DF, Bickel PJ, Hampel FR, Huber PJ, Rogers WH, Tukey JW (1972) Robust estimates of location: survey and advances. Princeton University Press, Princeton
Bertsimas D, Shioda R (2007) Classification and regression via integer optimization. Oper Res 55:252–271
Bickela DR, Frühwirthb R (2006) On a fast, robust estimator of the mode: comparisons to other robust estimators with applications. Comput Stat Data An 50:3500–3530
Cummins DJ, Andrews CW (1995) Iteratively reweighted partial least squares: a performance analysis by Monte Carlo simulation. J Chemometr 9:489–507
Dornheim H, Brazauskas V (2011) Robust-efficient fitting of mixed linear models: methodology and theory. J Stat Plan Infer 141:1422–1435
Hejazi TH, Bashiri M, Noghondarian K, Atkinson AC (2010) Multiresponse optimization with consideration of probabilistic covariates. Qual Reliab Eng Int 27:437–449
Huber PJ (1981) Robust statistics. Wiley, New York
Hund E, Massart DL, Smeyers-Verbeke J (2002) Robust regression and outlier detection in the evaluation of robustness tests with different experimental designs. Anal Chim Acta 463:53–73
Kazemzadeh RB, Bashiri M, Atkinson AC, Noorossana R (2008) A general framework for multiresponse optimization problems based on goal programming. Eur J Oper Res 189:421–429
Maronna RA, Martin RD, Yohai VJ (2006) Robust statistics: theory and methods. Wiley, New York
Massart DL, Kaufman L, Rousseeuw PJ, Leroy A (1986) Least median of squares: a robust method for outlier and model error detection in regression and calibration. Anal Chim Acta 187:171–179
Mendes B, Tyler DE (1995) Robust statistics, data analysis and computer intensive methods. Springer, New York
Morgenthaler S, Schumacher MM (1999) Robust analysis of a response surface design. Chemometr Intell Lab 47:127–141
Nguyena TD, Welsch R (2010) Outlier detection and least trimmed squares approximation using semi-definite programming. Comput Stat Data An 54:3212–3226
Ortiz MC, Sarabia LA, Herrero A (2006) Robust regression techniques. A useful alternative for detection of outlier data. Talanta 70:499–512
Pop HF, Sârbu C (1996) A new fuzzy regression algorithm. Anal Chem 68:771–778
Rousseeuw PJ (1984) Least median of squares regression. J Am Stat Assoc 79:871–880
Rousseeuw PJ, Yohai VJ (1984) Robust and nonlinear time series analysis. Springer, New York
Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York
Serneels S, Croux C, Filzmoser P, Van Espen PJ (2005) Partial robust M-regression. Chemometr Intell Lab 79:55–64
Shahriari H, Ahmadi O, Shokouhi AH (2011) A two-step robust estimation of the process mean using M-estimator. J Appl Stat 38:1289–1301
Siegel AF (1982) Robust regression using repeated medians. Biometrika 69:242–244
Zioutas G, Avramidis A (2005) Deleting outliers in robust regression with mixed integer programming. Acta Math Appl Sin 21:323–334

doi:10.1186/2251-712X-9-2
Cite this article as: Bashiri and Moslemi: **The analysis of residuals variation and outliers to obtain robust response surface.** *Journal of Industrial Engineering International* 2013 **9**:2.