Journal of Industrial Engineering International
a SpringerOpen Journal

ORIGINAL RESEARCH Open Access

Simultaneous robust estimation of multi-response surfaces in the presence of outliers

Mahdi Bashiri[1*] and Amir Moslemi[1,2]

Abstract

A robust approach should be considered when estimating regression coefficients in multi-response problems. Many models are derived from the least squares method. Because the presence of outlier data is unavoidable in most real cases and because the least squares method is sensitive to these types of points, robust regression approaches appear to be a more reliable and suitable method for addressing this problem. Additionally, in many problems, more than one response must be analyzed; thus, multi-response problems have more applications. The robust regression approach used in this paper is based on M-estimator methods. One of the most widely used weighting functions used in regression estimation is Huber's function. In multi-response surfaces, an individual estimation of each response can cause a problem in future deductions because of separate outlier detection schemes. To address this obstacle, a simultaneous independent multi-response iterative reweighting (SIMIR) approach is suggested. By presenting a coincident outlier index (COI) criterion while considering a realistic number of outliers in a multi-response problem, the performance of the proposed method is illustrated. Two well-known cases are presented as numerical examples from the literature. The results show that the proposed approach performs better than the classic estimation, and the proposed index shows efficiency of the proposed approach.

Keywords: Multi-response problem, Robust regression, Outliers, M-estimator

Introduction

A common method of explaining and analyzing the results of experiments is response surface modeling. 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 in each value of specific controllable factors. After gathering experimental data, a relationship between the factors (input data) and the response or responses (output results) should be defined to complete the analysis procedure. If we cannot construct a suitable model to define the precise relation between the input variables and the response or the responses' consequents, then the interpretations will not be reliable. After determining an experimental design and performing experiments, the next steps include the statistical analysis and the selection of the optimal input variables.

One of the most common approaches of regression coefficient estimation is the Least Squares (LS) method. A solution given by LS determines the coefficient values that minimize the sum of the squares of the residuals, in other words, the sum of the square differences between the experimental response values and those calculated by the fitted equation.

The quality of a manufactured product is often evaluated by several performance measures, which are called quality characteristics, each of which is described by a response variable. The values of these response variables are affected by one or more process parameters, which are the input variables. Often, processes with two or more response variables operate in a conflicting way. A group of responses often characterizes the performance of a manufactured product. These responses are usually correlated and measured by different measurement scales. Therefore, a decision-maker must resolve the parameter selection problem to optimize each response. This problem is considered to be a multi-response optimization problem, which is subject to different response requirements.

* Correspondence: bashiri@shahed.ac.ir
[1]Department of Industrial Engineering, Faculty of Engineering Shahed University, Tehran, Iran
Full list of author information is available at the end of the article

It is usually difficult to realize an optimal level of the input variables that can result in values close to the ideal or target values for all of the response variables. The main goal of multi-response optimization is, therefore, to find the settings of the input variables that achieve an optimal compromise in the response variables.

In many cases, especially in experimental results, some of the data should be treated outliers. These points, which may occur because of operator reading faults or other similar factors, may have a confusing effect on the total interpretation of the results. These points are called outliers. A data observation or a group of data points that are well separated from the majority of the whole pattern of observation, in other words, data that deviate from the general pattern, are called outliers. However, they are avoidable during the processing to some extent. The main concept in robustness is the presence of outliers and, more precisely, the changes in the distribution of the data. A common way to address outliers and to find them when using LS is to identify the bad observations. To detect the outliers, some graphical procedures such as normal probability plots and numerical regression diagnostics have been proposed. These procedures are defined in Weisberg (1985). Wisnowskia et al. (2001) studied the analysis of multiple outlier detection procedures for a linear regression model. Monte Carlo simulation is used to compare different approaches, and the performances and limitations of each method are discussed. Outlier detection in multivariate problems is not simple to understand; to describe this problem, simple visual methods can be applied. Fernandez Pierna et al. (2002) compared this type of method, called the convex hull method, with classical techniques and robust methods.

The concept of outlier data is qualitative in the sense that it is not the same as incorrect data but rather refers to data that are different from the majority. Often, the presence of outlier data illustrates the existence of an unexpected phenomenon at the start of experimentation but that can be explained, possibly from experimental causes. A problem that we often encountered in the application of regression is the presence of an outlier or outliers in the data. Outliers can be generated by a simple operational mistake, a small sample size, or other factors. Even one outlying observation can destroy an LS estimation, resulting in parameter estimates that do not provide useful information for the majority of the data. Robust regression analysis was developed to improve LS estimation in the presence of outliers and to provide additional information about valid observations. The primary purpose of robust regression analysis is to fit a model that represents the information that is in the majority of the data.

To address this obstacle, some robust approaches were proposed by different authors. Robust regression methods were introduced to address the above-mentioned problems. Ample (Fernandez Pierna et al. 2002) introduced robustness and computational approaches that include Huber (1981) robust statistics and different estimation algorithms. One common robust estimation approach is the M-estimator, which is based on a maximum likelihood estimation (MLE). LS was derived by this type of estimation and considers a special residuals function. The main idea of M-

Table 1 A brief review of single and multi-response robust regression in the literature

	Robust single response (using M-estimators)	Robust multi-response		
		Independent responses		Dependent responses (robust multivariate)
		Individuals	Simultaneous	
Hampel (1971)	✓			
Huber (1981)	✓			
Cummins and Andrews (1995)	✓			
Morgenthaler and Schumacher (1999)	✓			
Hund et al. (2002)	✓			
Wiens and Wu (2010)	✓			
Koksoy (2006)		✓		
Koksoy (2008)		✓		
Quesada and Del Castillo (2004)				✓
Daszykowski et al. (2007)				✓
Rousseeuw et al. (2004)				✓
Our research			✓	

Table 2 Experimental data of the tire tread compound problem

Experiment number	x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	−1	−1	+1	102	900	470	67.5
2	+1	−1	−1	120	860	410	65
3	−1	+1	−1	117	800	570	77.5
4	+1	+1	+1	198	2,294	240	74.5
5	−1	−1	−1	103	490	640	62.5
6	+1	−1	+1	132	1,289	270	67
7	−1	+1	+1	132	1,270	410	78
8	+1	+1	−1	139	1,090	380	70
9	−1.633	0	0	102	770	590	76
10	+1.633	0	0	154	1,690	260	70
11	0	−1.633	0	96	700	520	63
12	0	+1.633	0	163	1,540	380	75
13	0	0	−1.633	116	2,184	520	65
14	0	0	+1.633	153	1,784	290	71
15	0	0	0	133	1,300	380	70
16	0	0	0	133	1,300	380	68.5
17	0	0	0	140	1,145	430	68
18	0	0	0	142	1,090	430	68
19	0	0	0	145	1,260	390	69
20	0	0	0	142	1,344	390	70

estimators is to replace the squared residuals by another function. The M-estimator works by an iterative procedure. As a consequence, several authors (e.g., Cummins and Andrews 1995) have called this estimator iteratively reweighted least squares, or the IRLS method. Additionally in our case, to estimate the regression coefficients, the iterative weighting method can be applied to estimate robust coefficients. One M-estimator function is from Huber (1981), which has become increasingly popular. Since then, more robust approaches have been discussed by investigators. However since M-estimators are simple to understand and considering that recent methods are sometimes so sensitive that they wrongly identify good points as outliers (Hund et al. 2002), we chose to use these type of estimators. Maronna et al. (2006) explained the most recent robust regression algorithms.

Table 3 Summary of the least squares regression coefficients for each response in the first example

Coefficients	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$
x_1	16.49	268.15	−99.67	−1.41
x_2	17.88	246.5	−31.4	4.32
x_3	10.91	139.48	−73.92	1.63
x_1^2	−4.01	−83.55	7.93	1.56
x_2^2	−3.45	−124.79	17.31	0.06
x_3^2	−1.57	199.17	0.43	−0.32
x_1x_2	5.13	69.38	8.75	−1.63
x_1x_3	7.13	94.13	6.25	0.13
x_2x_3	7.88	104.38	1.25	−0.25
Intercept	139.12	1261.11	400.38	68.91

Morgenthaler and Schumacher (1999) discussed robust response surfaces in chemistry based on the design of experiments. Hund et al. (2002) presented various methods of outlier detection and evaluated robustness tests with different experimental designs. Robust regression methods and reconstruction experimental design methods have been compared. Wiens and Wu (2010) proposed a comparative study of M-estimators and presented a design that is more optimal compared with possible regression models.

In multi-response problems, the first step is the accurate determination of the regression coefficient because contamination and outlier data can have a negative effect on the models. The robustness concept in multi-responses has been presented by different authors; however, robust design was developed by Taguchi (1986, 1987). This approach is often used in process improvement project, to redesign processes for the purpose of increasing customer satisfaction by improving operational performances. Usually, the model parameters are estimated by LS in robust design. This methodology specifically utilizes both experimentation and optimization methods to determine the system's optimum operating conditions. Koksoy (2008, 2006) presented MSE as a robust design criterion in multi-response problems. Additionally, genetic algorithms and generalized reduced gradients method were used in their solution stage. In the mentioned studies, the general framework for multivariate problems in which data are collected from a combined array has been presented. For example, Quesada and Del Castillo (2004) proposed a dual response approach to multivariate robust parameter designs.

There are also several papers that consider correlations between responses, allowing the variance-covariance structure of the multiple responses to be accounted for. In this case, some multivariate techniques can be applied to these problems. Daszykowski et al. (2007) reviewed robust models and both univariate and multivariate outliers, and the effects of data analysis have been studied. One of the most efficient and useful robust multivariate regressions is the minimum covariance determinant, which was proposed by Rousseeuw et al. (2004).

In multi-response problems, robust regression approaches can be used to decrease the effects of contaminations and to focus outliers. In this paper, it is assumed that there is no correlation between responses;

Table 4 Scaled residuals of responses in the first iteration for the tire tread compound problem

Experiment number	R1	R2	R3	R4	Sum of absolute value of residuals
1	1.284007	0.604262	−0.97866	0.050079	2.917006
2	**1.608107**	−0.85258	−0.53883	0.674586	3.674095
3	0.563234	−0.83752	0.372694	0.891048	2.664497
4	0.612035	0.476129	0.181098	1.11425	2.383513
5	−0.47289	**−1.4899**	−0.49364	−1.05129	3.507725
6	−0.42409	−0.17625	−0.68524	−0.82808	2.113666
7	**−1.46897**	−0.1612	0.226282	−0.61162	2.468066
8	−1.14487	**−1.61804**	0.666114	0.012886	3.441902
9	0.123999	0.67264	0.386427	0.471895	1.654961
10	−0.33271	0.848008	0.082396	−0.56633	1.829443
11	−1.1557	0.691078	**1.502795**	0.737001	**4.086571**
12	0.946989	0.82957	−1.03398	−0.83144	3.641976
13	−0.27294	**2.456887**	−0.15226	−0.29297	3.175049
14	0.064229	−0.93624	0.621073	0.198524	1.820065
15	**−1.4959**	0.154146	−1.38046	0.811253	3.841755
16	**−1.4959**	0.154146	−1.38046	−0.30476	3.335263
17	0.215309	−0.46058	**2.005573**	−0.67677	3.35823
18	0.704225	−0.67871	**2.005573**	−0.67677	**4.065276**
19	**1.4376**	−0.00449	−0.70325	0.067244	2.212588
20	0.704225	0.32865	−0.70325	0.811253	2.547379

a similar assumption is made in other studies, such as Koksoy (2008). However, by considering each response and using univariate M-estimators to estimate the coefficients, a problem could occur, and outlier detection is required to consider all of the responses simultaneously. An outlier appearance in only one response cannot be considered a wrong observation while the other responses are considered to contain normal behavior. If we consider the responses individually, an experiment could treat an observation as outlier data, while in an iterative procedure, that point would be down-weighted more than it is appropriate. Considering all of the responses simultaneously leads to calculating the real number of outliers in a multi-response problem. In this paper, we suppose that the outlier data will occur because of a mistake in the experimentation, but not in the recording of the data. Hence considering one response as an outlier while considering others not to be

Table 5 Actual, individual robust, and SIMIR regression coefficient estimations for the tire tread compound problem

	$\hat{y}_1$			$\hat{y}_2$			$\hat{y}_3$			$\hat{y}_4$		
	Actual	Robust individual	SIMIR	Actual	Robust individual	SIMIR	Actual	Robust individual	SIMIR	Actual	Robust individual	SIMIR
x_1	14.77	17.17	16.75	280.27	302.80	276.86	−62.32	−89.92	−89.69	−1.41	−1.35	−1.57
x_2	19.59	17.79	17.13	258.6	248.54	242.62	−23.21	−31.37	−33.70	4.32	3.97	3.82
x_3	12.62	12.25	12.60	227.3	314.15	291.86	−45.45	−74.16	−75.05	1.63	1.68	1.42
x_1^2	−5	−4.52	−4.25	−10.95	−27.36	−23.64	5.24	12.05	7.18	1.56	1.54	1.64
x_2^2	−4.43	−3.85	−3.63	−52.2	−62.86	−53.89	7.33	17.42	16.68	0.06	0.11	0.16
x_3^2	−2.56	−2.07	−2.40	51.02	0.025	45.04	2.43	4.52	0.99	−0.32	−0.33	−0.13
x_1x_2	6.16	5.18	4.90	163.26	62.65	21.04	4.07	11.44	10.20	−1.63	−0.79	−0.57
x_1x_3	8.16	7.22	6.83	73.93	85.13	128.98	2.34	7.91	8.97	0.13	−0.21	0.14
x_2x_3	6.83	6.81	7.35	84.18	52.89	62.07	0.58	3.94	2.27	−0.25	0.58	0.82
Intercept	141.33	139.16	139.32	1241.7	1265.49	1225.18	396.04	391.21	403.17	68.91	68.85	68.74

Table 6 SE of the estimation of each regression coefficient in the least squares, individual robust and SIMIR approaches

	$\hat{y}_1$			$\hat{y}_2$			$\hat{y}_3$			$\hat{y}_4$		
	LS	Robust individual	SIMIR	LS	Robust individual	SIMIR	LS	Robust individual	SIMIR	LS	Robust individual	SIMIR
x_1	2.9584	5.79	3.93	146.89	507.94	11.61	1,395.02	762.00	749.18	0	0.0036	0.025
x_2	2.9241	3.20	6.05	146.41	101.17	255.27	67.07	66.62	110.22	0	0.12	0.25
x_3	2.9241	0.13	0.00	7,712.35	7,544.56	4,168.97	810.54	824.36	876.42	0	0.002	0.044
x_1^2	0.9801	0.22	0.55	5,270.76	269.58	161.11	7.23	46.46	3.79	0	0.0004	0.0064
x_2^2	0.9604	0.33	0.63	5,269.30	113.79	2.88	99.60	101.91	87.56	0	0.0025	0.01
x_3^2	0.9801	0.23	0.02	21,948.42	2,600.46	35.72	4	4.40	2.05	0	0.0001	0.036
x_1x_2	1.0609	0.95	1.57	8,813.45	10,122.05	20,225.64	21.90	54.37	37.59	0	0.70	1.12
x_1x_3	1.0609	0.88	1.75	408.04	125.47	3,030.72	15.28	31.05	43.96	0	0.115	0.0001
x_2x_3	1.1025	0.00	0.27	408.04	978.48	488.47	0.44	11.31	2.87	0	0.688	1.14
Intercept	4.8841	4.67	4.03	376.74	566.38	272.60	18.83	23.30	50.97	0	0.003	0.028
SSE	19.8356	16.43	18.83	50,500.43	22,929.92	28,653.03	2,439.95	1,925.83	1,964.64	0	1.645	2.66

SSE, sum of squared error.

contaminated is not rational. A brief review on the literature is given in Table 1, as follows:

To the best of our knowledge, there are a few studies on multi-response robust regression, and this paper focuses on the multi-response robust regression that considers the response residuals simultaneously. To estimate the regression coefficients in the multi-response problem, we propose a procedure in which we apply a simultaneous independent multi-response iterative reweighting procedure and change the M-estimator weighting function; a more precise estimation of each response can be obtained. By considering this procedure, a new criterion is proposed named the coincident outlier index (*COI*), and the performance of this procedure is analyzed.

This paper is organized as follows. 'Using M-estimators for robust estimation of regression coefficients' section presents the robust M-estimator procedure and the modification of the response surface by an iterative weighting procedure. The proposed method for the multi-response problem is defined in 'Robust simultaneous estimation of multi-response problem' section. To illustrate the proposed method, a numerical example is presented before the 'Conclusions' section. Finally, the last section provides the conclusions of this paper.

Using M-estimators for robust estimation of regression coefficients

The M-estimator proposed by Huber (1981) is the generalized form of the (MLEs). This part is extracted from

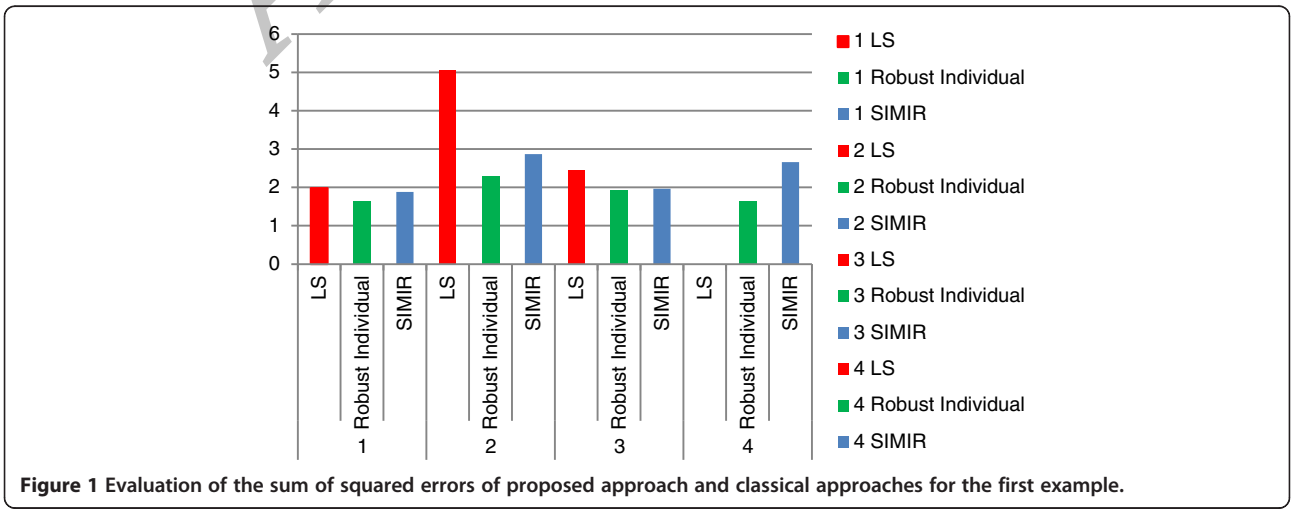

Figure 1 Evaluation of the sum of squared errors of proposed approach and classical approaches for the first example.

Maronna et al. (2006). The M-estimator is the solution of Equation (1), as follows:

$$\widehat{\theta} = \arg \min_{\theta} \sum_{i=1}^{n} \rho(x_i, \theta) \tag{1}$$

where ρ is a function with specific properties. Supposed that f is a density function and $\rho = -\log f$, where f is a density function, then $\widehat{\theta}$ will be introduced as the MLE of the parameter. There are several ρ functions. One common ρ function is Huber (1981). This function is well-defined in Equation (2):

$$\rho_k(x) = \begin{cases} x^2 & if \quad |x| \le k \\ 2k|x| - k^2 & if \quad |x| > k \end{cases} \tag{2}$$

By considering this function and by defining $\psi = \rho'$ and a weighting function by Equation (3), the iterative algorithm to estimate the unknown parameter can be defined.

$$W(x) = \begin{cases} \psi(x)/x & if \quad x \neq 0 \\ \psi'(x) & if \quad x = 0 \end{cases} \tag{3}$$

M-estimates for estimating regression coefficients are developed in the same way as defined in previous part. Equation (4) should be considered, and the coefficients can be obtained by solving following equation:

$$\sum_{i=1}^{n} \psi\left(\frac{r_i\left(\hat{\beta}\right)}{\hat{\sigma}}\right) \mathrm{x_i} = 0 \tag{4}$$

The $\widehat{\sigma}$ in Equation (4) can also be estimated individually by Equation (5), as follows, or can be solved simultaneously in Equation (4). r_i is a residual of the response.

$$\widehat{\sigma} = \frac{1}{0.675} \boldsymbol{Med}_i(|\boldsymbol{r}_i| \boldsymbol{r}_i \neq 0) \tag{5}$$

To make the estimation procedure invariant with respect to the scale of the residuals, the r_is are divided by 's'. 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

Table 7 Experimental data of the elastic element of the force transducer problem

Experiment number	x_1	x_2	x_3	z_1	z_2	y_1	y_2
1	−1	−1	−1	−1	1	1.810	1.10
2	−1	−1	−1	1	−1	1.690	1.11
3	−1	−1	1	−1	−1	1.900	1.07
4	−1	−1	1	1	1	1.780	1.07
5	−1	1	−1	−1	−1	1.800	1.47
6	−1	1	−1	1	1	1.630	1.18
7	−1	1	1	−1	1	1.920	1.41
8	−1	1	1	1	−1	1.780	1.58
9	1	−1	−1	−1	−1	1.360	1.57
10	1	−1	−1	1	1	1.220	2.03
11	1	−1	1	−1	1	1.480	1.38
12	1	−1	1	1	−1	1.440	1.68
13	1	1	−1	−1	1	0.693	3.37
14	1	1	−1	1	−1	0.616	3.75
15	1	1	1	−1	−1	0.950	2.81
16	1	1	1	1	1	0.817	2.83
17	−1	0	0	0	0	1.790	1.24
18	1	0	0	0	0	1.030	2.46
19	0	−1	0	0	0	1.530	1.23
20	0	1	0	0	0	1.220	1.73
21	0	0	−1	0	0	1.300	1.63
22	0	0	1	0	0	1.440	1.67
23	0	0	0	0	0	1.380	1.73
24	0	0	0	0	0	1.390	1.74
25	0	0	0	0	0	1.400	1.74

Table 8 Summary of the least squares regression coefficient for each response in the second example

Coefficients	$\hat{y}_1$	$\hat{y}_2$
x_1	−0.36	0.59
x_2	−0.15	0.43
x_3	0.07	−0.09
z_1	−0.05	0.06
z_2	−0.01	−0.04
$\boldsymbol{x}_1^2$	0.02	0.11
$\boldsymbol{x}_2^2$	0.2	−0.25
$\boldsymbol{x}_3^2$	0	−0.08
$\boldsymbol{z}_1^2$	0	0.33
x_1x_2	−0.14	0.3
x_1x_3	0.02	−0.14
x_1z_1	0.01	0.08
x_1z_2	0	0.01
x_2x_3	0.01	−0.033
x_2z_1	0	−0.03
x_2z_2	0	−0.06
x_3z_1	0	0
x_3z_2	0	−0.01
z_1z_2	0	−0.08
Intercept	1.39	1.73

bias adjustment for the standard deviation under the normal distribution.

An iterative reweighting method can be defined as follows: First, compute an initial estimate β_0 and compute $\widehat{\sigma}$ from Equation (5). After that, for $k = 0, 1, 2, \ldots$:

First, Compute an initial estimate β_0 and compute $\widehat{\sigma}$ from Equation (5). After that, For $k = 0, 1, 2, \ldots$:

(a) Given $\widehat{\beta}_k$, for $i = 1, \ldots, n$, compute $r_{i,k} = y_i - X_i'\widehat{\beta}_k$ and $w_{i,k} = W\left(r_{i,k}/\widehat{\sigma}\right)$.
(b) Compute $\widehat{\beta}_{k+1}$ by solving the following:

$$\sum_{i=1}^{n} w_{i,k} X_i \quad \left(y_i - X_i'\widehat{\beta}_k\right) = 0 \tag{6}$$

Finally, the algorithm stops when $\max\left(\left|r_{i,k} - r_{i,k+1}\right|\right)/\widehat{\sigma} < \varepsilon$

This algorithm converges if $W(x)$ is non-increasing for $x > 0$ (Maronna et al. 2006). If ψ is monotone, because the solution is essentially unique, the choice of the starting point influences the number of iterations but not the final result. This procedure is called (IRWLS).

The procedure is as follows: compute the first coefficients of the regression model, then compute the residuals and weights, and finally compute the new coefficients using Equation (6). This procedure can be repeated because the values of the coefficients and the values of the residuals and weights are different; as a result, this procedure can be repeated until a good solution is obtained. The procedure terminates when the change in the estimation from one iteration to the next is sufficiently small. The estimators of coefficients based on the LS method are basically unbiased; the robust estimators like LS methods are basically unbiased too.

Robust simultaneous estimation of multi-response problem

In a multi-response problem, similar to a single response problem, robust estimation of the regression model is an important issue. A simple approach to estimating the regression models in multi-response problems is to consider the responses individually and to estimate the solution to each problem by the robust M-estimator approach. However, this approach could cause some problems. Assume in one experiment that a specific response residual appears to be an outlier. However, other responses do not show any signs of being unacceptable data for that specific experiment. The outlier data could occur because of a fault in the experimentation. It is not rational to say, then, whether one experiment's result is an outlier for one response because it may not be unacceptable data for the other responses. From this type of deduction, a large amount of the experiment's results could become outliers, and for each response, some points will be assumed to be outliers by mistake. Thus, assuming independence between responses, a simultaneous independent multi-response iterative reweighting (SIMIR) approach is proposed to solve this problem. In this approach, based on M-estimators, some changes are applied in the procedure of weighting functions to estimate the coefficients of the model. The weighting function proposed in this method, down-weights the residuals by considering each response of the multi-response problem in each iteration, simultaneously. We have j responses in this problem and i experiments. The variable $r(i)_j$ defines the residual for the ith replicate of the jth response. The proposed weighting function is given in Equation (7):

$$\begin{cases} w_i = \dfrac{1}{c} & if \quad \forall \left|\boldsymbol{r}(\boldsymbol{i})_j\right| < c \\ w_i = \dfrac{1}{\left|\sum_{j=1}^{l} \boldsymbol{r}(\boldsymbol{i})_j\right|} & if \quad \exists \left|\boldsymbol{r}(\boldsymbol{i})_j\right| > c \end{cases}, \tag{7}$$

Table 9 Scaled residuals of two responses in the first iteration for proposed example

Experiment number	R1	R2	Sum of absolute value of residuals
1	−0.72	−0.172	0.894
2	−0.72	−0.172	0.894
3	−0.33	−0.892	1.229
4	−0.33	−0.892	1.229
5	−0.70	1.006	1.711
6	−0.70	1.006	1.711
7	−0.319	0.286	0.605
8	−0.319	0.286	0.605
9	0.319	−0.286	0.605
10	0.319	−0.286	0.605
11	0.704	−1.006	1.711
12	0.704	−1.006	1.711
13	0.337	0.892	1.229
14	0.337	0.892	1.229
15	0.722	0.172	0.894
16	0.722	0.172	0.894
17	**4.165**	−0.459	**4.624**
18	**−4.165**	0.459	**4.624**
19	0.072	**4.716**	**4.788**
20	−0.072	**−4.716**	**4.788**
21	**1.540**	**−2.880**	**4.420**
22	**−1.540**	**2.880**	**4.420**
23	**−2.166**	−0.166	2.333
24	0	0.083	0.083
25	**2.166**	0.083	2.250

The proposed pseudo code is as follows:

1. *Compute the actual values of the responses in each experiment by performing all of the experiments*
2. *Estimate the regression coefficients of the initial regression model by applying the proper method. While* $\left(\left|r_{i,k}-r_{i,k+1}\right|\right)/\widehat{\sigma} < \varepsilon$ *Do (ε is determined by the analyzer).*
3. *Calculate the residuals of each response in all of the experiments*
4. *Compute the* $\widehat{\sigma}$ *by Equation (5).*
5. *If all* $r(i)_j$ *are smaller than the threshold determined in the Huber M-estimator method, then dedicate the weight to be equal to 1 for the residuals in this iteration and go to step (7); else, go to step (6).*
6. *Down weight the residuals by considering the values of the residuals in all of the responses by a function in Equation (7).*
7. *Estimate the regression coefficients by solving Equation (6).*

The performance of the proposed method is presented by a numerical example in the next section. By applying the SIMIR procedure, the squared errors (SE) of the estimated parameters, which are regression coefficients, are reduced compared to those of the least squares estimation of each response; however, to some extent, this strategy is not as precise as the robust individual estimation. One important problem in multi-response problems is the number of real outliers.

The SE criterion is computed in Equation (8):

$$SE = \left(\theta - \hat{\theta}\right)^2 \tag{8}$$

Individually, residuals computed for one response could not be outliers in the whole multi-response problem. To detect the outliers in a multi-response problem, it is not correct to mention the outliers in each individual response. We present the COI to detect the real number of outliers in a robust estimation of the regression coefficients in a multi-response problem. This index can be computed by this procedure, for which we define the threshold by considering the suggested C (defined in Equation (7)) in the Huber procedure and by considering scaled residuals. If we consider the number of responses as n, then the proposed threshold is defined as $T = \left(\left\lceil \frac{n}{2} \right\rceil + 1\right) \times C$. If the sum of the residuals is greater than this threshold, then that experiment is treated as an outlier. Thus, the COI is equal to the number of points

Table 10 Actual, individual robust, and SIMIR approach for regression coefficient estimation for the proposed problem

	$\hat{y}_1$				$\hat{y}_2$			
	LS	**Actual**	**Robust individual**	**SIMIR**	**LS**	**Actual**	**robust Individual**	**SIMIR**
x_1	−0.36	−0.36	−0.35	−0.35	0.59	0.58	0.585	0.585
x_2	−0.15	−0.15	−0.15	−0.15	0.43	0.43	0.43	0.43
x_3	0.07	0.07	0.07	0.07	−0.09	−0.1	−0.09	−0.09
z_1	−0.05	−0.05	−0.05	−0.05	0.06	0.06	0.06	0.06
z_2	−0.01	−0.01	−0.01	−0.01	−0.04	−0.04	−0.04	−0.04
x_1^2	0.02	0.02	0.02	0.02	0.11	0.35	0.21	0.21
x_2^2	0.2	−0.01	0.09	0.09	−0.25	−0.2	−0.22	−0.22
x_3^2	0	−0.02	−0.01	−0.01	−0.08	0	−0.07	−0.07
z_1^2	0	0.05	0.02	0.02	0.33	0	0.15	0.15
x_1x_2	−0.14	−0.14	−0.14	−0.14	0.3	0.3	0.3	0.3
x_1x_3	0.02	0.02	0.02	0.02	−0.14	−0.14	−0.14	−0.14
x_1z_1	0.01	0.01	0.01	0.01	0.08	0.07	0.08	0.08
x_1z_2	0	0	0	0	0.01	0.01	0.01	0.01
x_2x_3	0.01	0.01	0.01	0.01	−0.033	−0.033	−0.033	−0.033
x_2z_1	0	0	0	0	−0.03	−0.03	−0.03	−0.03
x_2z_2	0	0	0	0	−0.06	−0.06	−0.06	−0.06
x_3z_1	0	0	0	0	0	0	0	0
x_3z_2	0	0	0	0	−0.01	0	0	0
z_1z_2	0	0	0	0	−0.08	0	0	0
Intercept	1.39	1.39	1.39	1.39	1.73	1.74	1.73	1.73

that are greater than T. Equation (9) defines the proposed COI, as follows:

$$\mathrm{COI} = \sum_{i=1}^{N} z_{ij} \qquad for \quad each \quad j = 1, \ldots, l$$

$$z_{ij} = \begin{cases} 1 & if \quad \sum_{j=1}^{4} r_{ij} > T \quad for \quad each \quad i \\ 0 & if \quad \sum_{j=1}^{4} r_{ij} < T \quad for \quad each \quad i \end{cases} \tag{9}$$

where N is the number of experiments.

Numerical example

In this section, the efficiency of our proposed approach compared with existing approaches is illustrated for two cases. In case one, the number of coinciding outliers differs from the number of outliers that were detected by an individual procedure. In the second case, the previous experiments contain most of the outliers because of a true fault by the experimenter. The number of coinciding outliers and the outliers that are detected individually does not differ significantly in this case.

Case 1. tire tread compound problem

In this section, we illustrate the proposed method using the well-known problem 'tire tread compound problem,' which was originally presented by Derringer and Suich (1980). In this model, three main chemical materials, such as silica (x_1), silane (x_2), and sulfur (x_3), and four responses are assumed. The experimental data results are given in Table 2.

As a first step, we attempt to find a primary regression model with four responses. A central composite design (CCD) with six center points is applied to describe the model. All of the controllable variables are $-1.63 \le x_i \le 1.63$, $i = 1, 2, 3$. The regression coefficients that are obtained by the least squares estimation method and according to the CCD are given in Table 3, as follows:

The scaled residuals of this multi-response problem are reported in Table 4, as follows:

For the first response, the residuals obtained by the experiments numbered 2, 7, 15, 16, and 19 appear to be outliers, and for the second response residuals, those numbered 5, 8, and 13 appear to be outliers. For the third

Table 11 SE of the estimation of regression coefficients in LS, robust individual, and SIMIR methods

	$\hat{y}_1$			$\hat{y}_2$		
	LS	**Robust individual**	**SIMIR**	**LS**	**robust Individual**	**SIMIR**
x_1	0.0001	0	0	0.0001	0.000025	0.000025
x_2	0	0	0	0	0	0
x_3	0	0	0	0.0001	0.0001	0.0001
z_1	0	0	0	0	0	0
z_2	0	0	0	0	0	0
x_1^2	0	0	0	0.057	0.019	0.019
x_2^2	0.0441	0.01	0.01	0.0025	0.0004	0.0004
x_3^2	0.0004	0.0001	0.0001	0.006	0.0049	0.004
z_1^2	0.0025	0.0009	0.0009	0.108	0.022	0.022
x_1x_2	0	0	0	0	0	0
x_1x_3	0	0	0	0	0	0
x_1z_1	0	0	0	0	0.0001	0.0001
x_1z_2	0	0	0	0	0	0
x_2x_3	0	0	0	0	0	0
x_2z_1	0	0	0	0	0	0
x_2z_2	0	0	0	0	0	0
x_3z_1	0	0	0	0	0	0
x_3z_2	0	0	0	0.0001	0	0
z_1z_2	0	0	0	0.0064	0	0
Intercept	0	0	0	0.0001	0.0001	0.0001
SSE	0.0471	0.011	0.011	0.1823	0.047	0.047

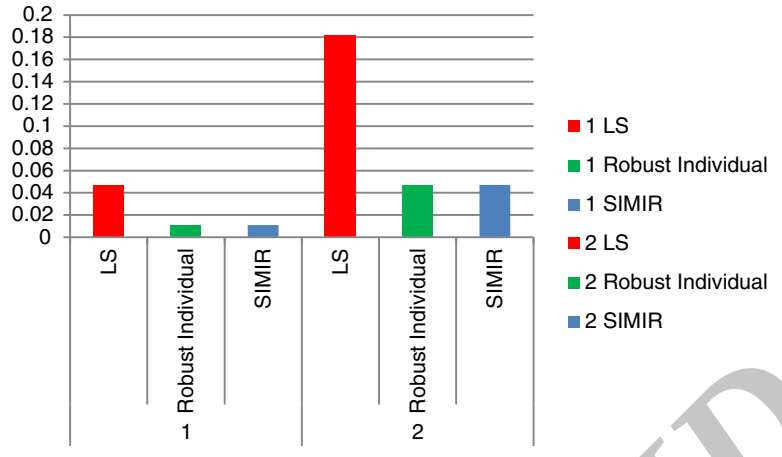

Figure 2 Evaluation of the sum of squared errors of proposed and classical approaches for the second example.

response residuals, those numbered 11, 17, and 18 appear to be outliers. The fourth model does not contain a residual that implies an outlier. By omitting the values of the implied outliers, actual regression models can be obtained by a least squares method because least square is the most efficient method in the absence of outlier data. This method is considered to be the actual result in Table 5. In addition, to obtain the robust regression models by two different approaches, the robust individual regression approach and the SIMIR approach are applied and can be compared by considering the SE criteria.

First, we consider each response individually and apply the M-estimator procedure to each response. The constant C in this example is assumed to be 1.37. Finally, the SIMIR procedure is applied to the data. Coefficients obtained by actual, robust individual and SIMIR procedures are given in Table 5.

Then, to evaluate the procedures mentioned (actual, individual robust, and SIMIR approaches), the SE of these estimated parameters are computed using Equation (8), and the results are reported in Table 6. In Equation (8), it is assumed that the value of θ is the regression coefficient in the actual method, and $\hat{\theta}$ is the regression coefficient in the considered method (actual, individual robust, and SIMIR approaches).

Additionally, the evaluation is given in Figure 1, as follows:

A comparison among the three approaches by considering the sum of squared errors (SSE) is computed in Table 6 and illustrated in Figure 1. In this figure, for the first three responses, the scaled values of the SSE are given.

The results show that the robust individual regression estimation is more precise than that of the least squares estimation or the proposed SIMIR approach, but the coincident outlier index that was presented in the previous section is more reliable and realistic. Moreover, the results state that in the case with an absence of outliers, LS performs better than both the robust individual and the SIMIR procedures. In this example, if we want to count the outliers as a multi-response problem individually, the results would be 11 experiments. However, by the proposed COI index, the real number of the multi-response problem outliers is 2, and it would be more rational that in 20 experiments, almost 10% of the experiments result in outliers.

Case 2. elastic element of a force transducer problem

We provide another example to illustrate the efficiency of the proposed method. The following example was presented as a case study in Romano et al. (2004), in which the problem was about the elastic element of a force transducer. This example involves a combined array design with three control (x) and two noise (z) variables. The control factors are the three parameters that describe the element configuration, namely the lozenge angle (x1), the bore diameter (x2), and the half-length of the vertical segment (x3). Noise factors are the deviation of the lozenge angle from its nominal value (z1) and the

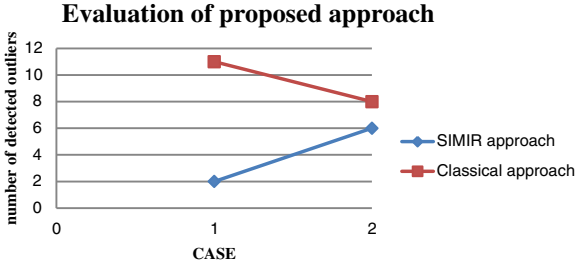

Figure 3 Evaluating the performance of the proposed method and the existing method for both cases, by considering the COI index.

deviation of the bore diameter from its nominal value (z2). These internal noise factors are undeniably independent. The two indicators, namely the non-linearity (y1) and the hysteresis (y2), define the responses. Table 7 displays the data from this experiment.

First, we need to find a primary regression model for the two responses. All of the controllable variables are $-1 \le x_i \le 1, i = 1, 2$. The regression coefficients that were obtained by the least squares estimation method are given in Table 8, as follows:

The scaled residuals of this multi-response problem are reported in Table 9, as follows:

Consequently, for the first response, outliers appear with the residuals obtained by the experiments, numbered 17, 18, 21, 22, 23 and 25; for second response residuals, they are numbered 19, 20, 21 and 22. By omitting these values, the actual regression models can be obtained. To obtain the robust regression models, two different approaches are applied, and these two approaches are compared by considering the SE criteria.

First, we consider each response individually and apply the M-estimator procedure to each response. The constant C in this example is assumed to be 1.37. Thus, three groups of coefficients are reported in Table 10.

To evaluate the three procedures, the proposed SE criterion are calculated, and the results are given in Table 11 as follows:

In addition, to provide more illustration, Figure 2 is given as follows:

Similar to the previous example, our results showed that the robust individual regression estimation performs better, but not significantly better than the least squares method; the SIMIR approach was also considered, but the COI is more realistic and accurate. In this example, if we want to count the outliers in a multi-response problem individually, the results would consist of eight experiments. However, by the proposed COI index, the real number of outliers in the multi-response problem is six.

Therefore, by considering these two examples, the number of detected outliers calculated by both the classical method and the SIMIR proposed approach is shown in Figure 3, as follows:

Conclusions

As mentioned in the previous sections, a robust simultaneous estimation of regression coefficients in the multi-response problem in the case in which contaminated data exists was presented in this paper. In addition, the results showed that the proposed approach would consider a number of points to be real outliers in the multi-response problem, although individual robust regression shows some other points as outliers. Thus, an aggregative approach in the weighting function was proposed, in which all of the responses were surveyed. The SIMIR approach performed better than the classic method for detecting outliers and estimating regression coefficients. Additionally, our results show that the proposed approach would provide a better COI index than the classical approach for outlier detection. For future research, other robust regression approaches can be studied. In addition, considering a problem with correlated responses can be another aspect of related future research.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MB has worked on the literature and modeling; he also proposed the response surface methodology. AM has performed the robust concept and numerical example. Both authors read and approved the final manuscript.

Authors' information
MB is an associate professor of the Department of Industrial Engineering, Shahed University, and his research interests are Multiple Response Optimization and Facilities Location Problem. AM is a Ph.D. student of the Department of Industrial Engineering at Amirkabir University of Technology (Tehran Polytechnic).

Acknowledgments
Authors are thankful to the reviewers for their valuable comments.

Author details
[1]Department of Industrial Engineering, Faculty of Engineering Shahed University, Tehran, Iran. [2]Department of Industrial Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez, Tehran, Iran.

Received: 15 March 2013 Accepted: 19 March 2013
Published: 17 April 2013

References

Cummins DJ, Andrews CW (1995) Iteratively reweighted partial least squares: a performance analysis by Monte Carlo simulation. J Chemometr 9:489–507
Daszykowski M, Kaczmarek K, Vander Heyden Y, Walczak B (2007) Robust statistics in data analysis–a review basic concepts. Chemometr Intell Lab 85:203–219
Derringer G, Suich R (1980) Simultaneous optimization of several response variables. J Qual Technol 12:214–219
Fernandez Pierna JA, Wahl F, De Noord OE, Massart DL (2002) Methods for outlier detection in prediction. Chemometr Intell Lab 63:27–39
Hampel FR (1971) A general definition of qualitative robustness. Ann Math Statist 42:1887–1896
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
Huber PJ (1981) Robust statistics. Wiley, New York
Maronna RA, Martin RD, Yohai VJ (2006) Robust statistics: theory and methods. Wiley, New York
Morgenthaler S, Schumacher MM (1999) Robust analysis of a response surface design. Chemometr Intell Lab 47:127–141
Koksoy O (2008) A nonlinear programming solution to robust multiresponse quality problem. Appl Math Comput 196:603–612
Koksoy O (2006) Multiresponse robust design: mean square error (MSE) criterion. Appl Math Comput 175:1716–1729
Quesada GM, Del Castillo E (2004) A dual-response approach to the multivariate robust parameter design problem. Technometrics 46:176–187
Romano D, Varetto M, Vicario G (2004) Multiresponse robust design: a general framework based on combined array. J Qual Technol 36:27–37
Rousseeuw P, Van Aelst S, Van Driessen K, Agull J (2004) Robust Multivariate Regression. Technometrics 46:293–305

Taguchi G (1986) Introduction to quality engineering. Kraus International Publications, White Plains, NJ
Taguchi G (1987) System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost. Quality Resources, White Plains, NJ
Weisberg S (1985) Applied linear regression, 2nd edition. Wiley, New York
Wiens D, Wu EKH (2010) A comparative study of robust designs for M-estimated regression models. Comput Stat Data An 54:1683–1695
Wisnowskia JW, Montgomery DC, Simpson JR (2001) A comparative analysis of multiple outlier detection procedures in the linear regression model. Comput Stat Data An 36:351–382

doi:10.1186/2251-712X-9-7
Cite this article as: Bashiri and Moslemi: **Simultaneous robust estimation of multi-response surfaces in the presence of outliers.** *Journal of Industrial Engineering International* 2013 **9**:7.