ISTANBUL UNIVERSITY PRESS

Istanbul Journal of Mathematics
ijmath 2 (2), 95–108, 2024
DOI: 10.26650/ijmath.2024.00020

RESEARCH ARTICLE

A New Liu-Ratio Estimator For Linear Regression Models

İ. M. Giresunlu[1], K. U. Akay[1*], and E. Ertan[1]

[1]İstanbul University, Faculty of Science, Department of Mathematics, Vezneciler, 34134, İstanbul, Türkiye

ABSTRACT
In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. Although there are various methods for estimating parameters, the most popular is the Ordinary Least Squares (OLS) method. However, in the presence of multicollinearity and outliers, the OLS estimator may give inaccurate values and also misleading inference results. There are many modified biased robust estimators for the simultaneous occurrence of outliers and multicollinearity in the data. In this paper, a new estimator called the Liu-Ratio Estimator (LRE), which can be used as an alternative to the Least Squares Ratio (LSR) estimator and the Ridge Ratio estimator (RRE), is proposed to mitigate the effect of y-direction outliers and multicollinearity in the data. The performance of the proposed estimator is examined in two Monte Carlo simulation studies in the presence of multicollinearity and y-direction outliers. According to the simulation results, LRE is a strong alternative to LSR and RRE in the presence of multicollinearity and y-direction outliers in the data.

Mathematics Subject Classification (2020): 62J05,62J07

Keywords: Least Squares Ratio Estimator, Liu Estimator, Multicollinearity, Ridge Estimator.

1. INTRODUCTION

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. Applications for regression models are numerous and occur in almost every field, including engineering, the physical and chemical sciences, economics, management, life and biological sciences, and social sciences. The classical linear regression model assumes a relation of the form:

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i, \quad i = 1, 2, ..., n \tag{1}$$

where n is the number of observations, x_{ij} $j = 1, 2, ..., p$ are the independent variables for observation i, y_i the observed response variable, the ε_i is the error term for the observation i and β_j are the coefficients to be estimated, representing the relationship between each independent variable and the dependent variable.

The most popular way of estimating β is to minimize the Ordinary Least Squares (OLS) criterion. Unfortunately, the well-known problem of multicollinearity in regression analysis due to high correlation between independent variables affects the OLS estimator. As a result of multicollinearity between explanatory variables, the variance of OLS becomes so large that estimates become unstable (Montgomery et al. 2001). Many biased estimators have been proposed for the multicollinearity problem, but the Ridge Estimator (RE) proposed by Hoerl and Kennard (1970) and the Liu Estimator (LE) proposed by Liu (1993) are some of the most widely used estimators.

In addition, there are many situations where the distribution of errors is nonnormal. In the case of nonnormal distributions, particularly heavy-tailed distributions, the OLS estimator no longer has the desirable properties. These heavy-tailed distributions tend to generate outliers, which may have an improper effect on the OLS estimates (Montgomery et al. 2001). Numerous robust estimating techniques, including the M-estimator, the least squares median estimator, the least truncated sum of squares estimator, the S-estimator, and the MM-estimator, have been presented to generate parameter estimates in the presence of outliers (Rousseeuw and Leroy 1987), (Maronna et al. 2006). However, while robust estimators are robust techniques for obtaining parameter estimates that are not affected by outliers, some unstable estimates may still be obtained due to the presence of multicollinearity between variables. Therefore, to mitigate the effects of both outliers and multicollinearity to some extent is to use biased-robust estimators.

Corresponding Author: K. U. Akay **E-mail:** kulas@istanbul.edu.tr
Submitted: 17.09.2024 • **Revision Requested:** 27.12.2024 • **Last Revision Received:** 30.12.2024 • **Accepted:** 31.12.2024

For example, various modifications of RE and LE, which are used for the multicollinearity problem, are widely used to address both outliers and multicollinearity (Silvapulle 1991), (Arslan and Billor 2000), (Maronna 2011), (Kan et al. 2013), (Jadhav and Kashid 2016), (Ertaş et al. 2017), (Filzmoser and Kurnaz 2018).

Recently, Akbilgic and Akinci (2009) proposed the Least Squares Ratio (LSR) as an alternative for OLS in order to estimate the beta parameter vector in the presence of y-direction outliers. On the other hand, Jadhav and Kashid (2018) developed an estimator called the Ridge Ratio Estimator (RRE) as an alternative to RE and LSR in the presence of outliers and multicollinearity in the data. Therefore, one of the objectives of this paper is to propose a new estimator as an alternative to LSR and RRE to overcome the simultaneous occurrence of outliers and multicollinearity in the data, based on the fact that LE is always an alternative to RE as known from the multicollinearity problem. Another objective is to investigate the performance of the proposed estimator with respect to LSR and RRE through extensive simulation studies.

The organization of the paper is as follows: The main ideas underlying the proposed estimator are highlighted in Section 2. In Section 3, two separate Monte Carlo simulation studies are conducted to evaluate the performance of the proposed estimator with respect to LSR and RRE. In Section 4, the performance of the proposed estimator is evaluated against that of other estimators on artificial data. Finally, the conclusions of the study are presented in Section 5.

2. A NEW ROBUST LIU RATIO ESTIMATOR

For the regression model given by (1), OLS minimizes the sum of squares of the distances between the observed value y_i and the fitted value $\hat{y}_i$ where $i = 1, 2, ..., n$. As an alternative to OLS, LSR method starts with the same goal $y_i = \hat{y}_i$, or $y_i - \hat{y}_i = 0, \quad i = 1, 2, ..., n$ as in OLS. Note that the OLS approach satisfies this aim by finding the regression parameters minimizing the sum of $(y_i - \hat{y}_i)^2$. However, LSR proceeds by dividing through by y_i and so $\frac{\hat{y}_i}{y_i} = 1$ is obtained under an assumption of $y_i \neq 0$ where $i = 1, 2, ..., n$ (Akbilgic and Akinci 2009). Hence, it is obvious that, equations $\frac{\hat{y}_i}{y_i} - 1 = 0$ and thus $\frac{y_i - \hat{y}_i}{y_i} = 0$ where $i = 1, 2, ..., n$ are obtained by basic mathematical operations. As a result, the LSR estimator is obtained by minimizing the objective function as follows:

$$\min_{\beta} \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \quad \text{or} \quad \min_{\beta} \sum_{i=1}^{n} \left(1 - \hat{\beta}_j \frac{x_{ij}}{y_i} \right)^2 \tag{2}$$

where $\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}, \quad i = 1, 2, ..., n$. Taking the partial derivatives of (2) with respect to the β components and setting them equal to zero, Akbilgic and Akinci (2009) defined the LSR estimator as follows:

$$\hat{\beta}_{LSR} = \left(\left(\frac{X}{Y} \right)' \left(\frac{X}{Y} \right) \right)^{-1} \left(\frac{X}{Y^2} \right)' Y \tag{3}$$

where X/Y matrix is obtained by dividing the values x_{ij} by y_i, and X/Y^2 is computed by dividing the values x_{ij} by y_i^2 where $j = 1, 2, ..., p$.

On the other hand, Jadhav and Kashid (2018) developed an estimator called RRE as an alternative to RE and LSR. Note that RRE using RE and LSR estimator is proposed to tackle the problem of outliers and multicollinearity. For the parameters β in Equation (1), the RRE is defined as:

$$\hat{\beta}_{RRE} = \left(\left(\frac{X}{Y} \right)' \left(\frac{X}{Y} \right) + kI \right)^{-1} \left(\frac{X}{Y^2} \right)' Y, \quad k > 0, \tag{4}$$

where k is a biasing parameter.

Let us state that the LSR and RRE given by (3) and (4) are obtained by minimization of the objective function given below:

$$S(\beta) = (1 - \underline{X}\beta)' (1 - \underline{X}\beta) + k\beta'\beta \tag{5}$$

where 1 is the $n \times 1$ dimensional matrix of 1s, $\underline{X}$ is obtained by dividing the values x_{ij} by y_i for $j = 1, ..., p$ and the parameter $k \geq 0$ controls the amount of shrinkage. Note that minimization of the objective function given by (5) with respect to the parameter vector β yields the LSR estimator given by (3) when $k = 0$ and the RRE given by (4) when $k > 0$.

As an alternative to the objective function (5), which yields the LSR and RRE given by (3) and (4), consider the following penalized objective function:

$$S(\beta) = (1 - \underline{X}\beta)' (1 - \underline{X}\beta) + (d\hat{\beta}_{LSR} - \beta)' (d\hat{\beta}_{LSR} - \beta), \qquad 0 < d < 1 \tag{6}$$

where $\hat{\beta}_{LSR}$ is the LSR estimator given in (3) and $\underline{X}$ is obtained by dividing the values x_{ij} by y_i for $j = 1, ..., p$. When $S(\beta)$ in (6) is differentiated with respect to β, the following equation is obtained:

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\beta}} = -2\underline{X}' + 2\underline{X}'\underline{X}\beta - 2d\hat{\beta}_{LSR} + 2\beta = 0. \tag{7}$$

Solving the system given in (7) with respect to β defines the Liu Ratio Estimator (LRE) as follows:

$$\hat{\beta}_{LRE} = \left(\underline{X}'\underline{X} + I\right)^{-1}\left(\underline{X}' + d\hat{\beta}_{LSR}\right), \qquad 0 < d < 1, \tag{8}$$

where d is a biasing parameter. If the estimator (8) is restated in the structure of (3) or (4), LRE is obtained as follows:

$$\hat{\beta}_{LRE} = \left(\left(\frac{X}{Y}\right)'\left(\frac{X}{Y}\right) + I\right)^{-1}\left(\left(\frac{X}{Y^2}\right)' Y + d\hat{\beta}_{LSR}\right), \qquad 0 < d < 1 \tag{9}$$

where X/Y matrix is obtained by dividing the values x_{ij} by y_i, and X/Y^2 is computed by dividing the values x_{ij} by y_i^2 where $j = 1, 2, ..., p$.

3. THE MONTE CARLO SIMULATION STUDIES

In this section, the performance of LRE is compared with other existing estimators, OLS, RE, LE, LSR and RRE using two different Monte Carlo simulation designs. In the first design, we investigated the effects of sample size (n), the degree of the collinearity (ρ), the number of the explanatory variables (p) and the variance (σ^2) on the performances of the considered estimators. In the second simulation design, we examined LSR, RRE and LRE performances for each of n, p, ρ and σ^2 values at certain values of k and d. For both simulation designs, we generate the explanatory variables by the following McDonald and Galarneau (1975) as

$$x_{ij} = \left(1-\rho^2\right)^{1/2} u_{ij} + \rho u_{ip+1}, \quad i = 1, 2, .., n, \quad j = 1, 2, ..., p \tag{10}$$

where u_{ij} are independent standard normal pseudo-random numbers. ρ is specified so that the correlation between any two variables is given by ρ^2. These variables are standardized such that $X'X$ is a correlation matrix. Investigations are conducted on three distinct sets of correlations that correspond to $\rho = 0.8, 0.9$ and 0.95. The response variable is generated by

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{11}$$

where $\varepsilon_i \sim N\left(0, \sigma^2\right)$ and β_0 is equal to zero. The values of σ^2 are 1, 5, and 10 for various comparisons of the error term. For each set of explanatory variables, the parameter vector β is chosen as the normalized eigenvector corresponding to the largest eigenvalue of $X'X$ so that $\beta'\beta = 1$. The sample sizes n are 50, 100 and 200. The number of explanatory variables is chosen as $p = 4$, 8, and 12.

We examine the effects of y-direction outliers on the estimators by considering three different cases such as no outlier, one outlier and two outliers. When there is no outlier, dependent variables are taken into consideration as in Equation (11). In the case of one outlier, the n observation is changed as $y(n) = 500$. For two outlier case, $y(1) = 500$ and $y(n) = 500$ altered observations are used.

In order to estimate the biasing parameters in the simulation, based on the studies of Kibria (2003) and Qasim et al. (2020), the biasing parameters for RE, LE, RRE, and LRE are taken as follows:

RE: $\hat{k}_{RE} = \frac{\hat{\sigma}^2_{OLS}}{\left(\prod_{j=1}^{p+1}\hat{\beta}^2_{OLS(j)}\right)^{\frac{1}{p+1}}}$ where $\hat{\sigma}^2_{OLS} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{OLS(i)}\right)^2}{n-p-1}$

LE: $\hat{d}_{LE} = \max\left(0, \min\left(\frac{\hat{\beta}^2_{OLS(j)} - \hat{\sigma}^2_{OLS}}{\max\left(\frac{\hat{\sigma}^2_{OLS}}{\lambda_j}\right) + \max\left(\hat{\beta}^2_{OLS(j)}\right)}\right)\right)$ where λ_j is the jth eigenvalues of $X'X$, $j = 1, 2, ..., p+1$.

RRE: $\hat{k}_{RRE} = \frac{\hat{\sigma}^2_{LSR}}{\left(\prod_{j=1}^{p+1}\hat{\beta}^2_{LSR(j)}\right)^{\frac{1}{p+1}}}$ where $\hat{\sigma}^2_{LSR} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{LSR(i)}\right)^2}{n-p-1}$

LRE: $\hat{d}_{LRE} = \max\left(0, \min\left(\frac{\hat{\beta}^2_{LSR(j)} - \hat{\sigma}^2_{LSR}}{\max\left(\frac{\hat{\sigma}^2_{LSR}}{\lambda_j}\right) + \max\left(\hat{\beta}^2_{LSR(j)}\right)}\right)\right)$ where λ_j is the jth eigenvalues of $\underline{X}'\underline{X}$, $j = 1, 2, ..., p+1$.

As a measure of performance we use the estimated Mean Squared Error (MSE) between the estimated parameters in the l-th repetition, $\hat{\beta}^{(l)}$, and the true parameters β:

$$\text{MSE} = \frac{1}{m}\sum_{l=1}^{m}\frac{1}{p}\left\|\beta - \hat{\beta}^{(l)}\right\|^2 \tag{12}$$

where p is the number of explanatory variables. The simulation experiment is replicated $m = 2000$ times by generating new pseudo-random numbers. The R programming language was used to carry out the calculations. The results are given in Tables 1-3 where the lowest estimated MSE values in each row are indicated by bold.

In all 81 scenarios in Tables 1-3, the LSR, RRE and LRE outperformed other estimators according to criterion (12). With the

Table 1. The estimated MSE values of the considered estimators for the model when $p = 2$

			No outlier						One outlier						Two outliers					
σ^2	n	ρ	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**
1	50	0.8	3.874	1.731	**0.836**	1.104	1.008	1.043	66639.985	33649.654	10840.251	1.115	**1.011**	1.045	47482.474	13934.638	6376.756	1.119	**1.012**	1.048
5	50	0.8	18.756	7.000	3.602	1.796	1.243	**1.229**	66755.55	33738.42	10868.439	1.893	1.266	**1.242**	47585.798	13989.384	6389.664	1.906	1.27	**1.245**
10	50	0.8	39.375	14.953	8.504	2.547	1.466	**1.324**	66525.692	33542.163	10818.309	2.753	1.53	**1.345**	47403.547	13909.209	6374.627	2.781	1.534	**1.340**
1	50	0.9	8.586	3.144	1.367	1.33	**1.072**	1.102	10400.479	275.513	271.375	1.340	**1.076**	1.104	335131.132	160914.664	89564.179	1.392	**1.088**	1.114
5	50	0.9	42.251	14.49	7.867	3.02	1.605	**1.443**	10412.218	279.069	271.616	3.058	1.611	**1.444**	335134.314	160891.027	89610.788	3.364	1.699	**1.499**
10	50	0.9	81.964	27.041	14.999	4.822	2.097	**1.629**	10387.775	283.39	273.565	4.875	2.116	**1.637**	334751.036	160565.563	89372.943	5.368	2.259	**1.711**
1	50	0.95	14.332	4.986	2.313	1.565	**1.159**	1.178	110682.31	22309.457	16422.522	1.595	**1.167**	1.181	25303.971	310.034	239.406	1.619	**1.175**	1.188
5	50	0.95	72.727	25.501	15.425	4.03	1.88	**1.614**	110530.006	22269.075	16397.836	4.228	1.94	**1.646**	25272.394	313.56	245.276	4.375	1.981	**1.663**
10	50	0.95	142.487	46.331	28.765	6.875	2.787	**2.124**	110708.454	22374.722	16479.652	7.279	2.878	**2.203**	25391.9	321.936	253.227	7.340	2.865	**2.200**
1	100	0.8	3.34	1.471	**0.720**	1.036	0.996	1.02	10198.92	2144.602	984.846	1.038	**0.996**	1.021	480.365	3.225	85.917	1.042	**0.998**	1.024
5	100	0.8	17.474	6.768	3.864	1.396	**1.121**	1.169	10203.427	2151.653	999.082	1.404	**1.123**	1.171	496.307	3.431	87.961	1.421	**1.127**	1.175
10	100	0.8	34.438	12.756	7.631	1.771	1.233	**1.225**	10251.047	2176.151	1018.585	1.795	1.237	**1.228**	505.923	3.628	89.693	1.818	1.243	**1.233**
1	100	0.9	4.788	1.916	**0.851**	1.062	1.003	1.035	6251.16	362.017	264.557	1.063	**1.003**	1.036	1603.233	83.091	659.89	1.065	**1.004**	1.037
5	100	0.9	23.743	8.541	4.500	1.432	**1.109**	1.147	6287.381	372.356	270.937	1.442	**1.113**	1.151	1625.594	85.726	662.281	1.449	**1.114**	1.153
10	100	0.9	49.678	17.979	9.756	2.015	1.312	**1.281**	6314.094	384.904	279.665	2.021	1.31	**1.28**	1644.439	87.799	661.543	2.045	1.319	**1.286**
1	100	0.95	13.699	4.709	2.265	1.262	**1.058**	1.104	16134.448	1079.382	504.879	1.264	**1.059**	1.104	3786.59	659.29	1599.82	1.272	**1.062**	1.107
5	100	0.95	71.713	23.493	14.264	2.583	1.463	**1.378**	16206.988	1099.17	516.107	2.598	1.462	**1.378**	3838.046	669.735	1598.714	2.615	1.470	**1.382**
10	100	0.95	129.272	41.231	24.767	3.997	1.865	**1.656**	16270.935	1124.121	529.387	4.054	1.879	**1.663**	3910.352	688.784	1606.804	4.083	1.886	**1.662**
1	200	0.8	3.846	1.639	**0.774**	1.01	0.993	1.006	7926.647	1957.891	1343.945	1.010	**0.993**	1.006	5383.217	563.529	526.902	1.010	**0.993**	1.006
5	200	0.8	18.474	6.78	3.755	1.203	**1.062**	1.114	7965.217	1981.025	1363.837	1.206	**1.064**	1.115	5415.279	572.762	529.15	1.207	**1.063**	1.116
10	200	0.8	39.446	14.739	8.691	1.406	**1.125**	1.176	7985.94	1998.333	1381.326	1.420	**1.130**	1.18	5435.076	581.752	530.902	1.419	**1.129**	1.178
1	200	0.9	5.757	2.24	**0.964**	1.04	1.003	1.029	1723.361	483.201	632.332	1.040	**1.003**	1.029	14523.593	4669.841	2203.649	1.040	**1.002**	1.029
5	200	0.9	29.047	10.274	5.329	1.276	**1.072**	1.124	1741.125	488.224	633.732	1.279	**1.073**	1.125	14504.736	4667.649	2204.356	1.285	**1.076**	1.127
10	200	0.9	56.529	19.381	10.749	1.645	**1.185**	1.219	1771.988	495.737	634.393	1.651	**1.188**	1.221	14577.619	4691.152	2204.976	1.652	**1.189**	1.223
1	200	0.95	10.419	3.527	1.636	1.096	**1.018**	1.058	40589.394	15480.793	8489.903	1.099	**1.02**	1.06	38599.933	8068.549	3190.09	1.099	**1.020**	1.059
5	200	0.95	54.636	17.793	10.267	1.58	**1.154**	1.197	40580.392	15466.929	8503.168	1.595	**1.157**	1.199	38592.549	8082.718	3204.612	1.604	**1.164**	1.202
10	200	0.95	113.452	38.663	23.827	2.237	1.359	**1.337**	40655.728	15509.057	8562.385	2.299	1.377	**1.346**	38665.172	8132.565	3237.893	2.297	1.376	**1.344**

Table 2. The estimated MSE values of the considered estimators for the model when $p = 4$

			No outlier						One outlier						Two outliers					
σ^2	n	ρ	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**
1	50	0.8	8.284	2.658	**0.908**	1.471	1.064	1.21	44899.659	19414.94	12989.991	1.490	**1.069**	1.217	69474.34	18211.508	10069.709	1.505	**1.073**	1.220
5	50	0.8	42.778	12.939	4.247	3.907	1.693	**1.481**	44931.556	19423.137	12921.074	4.051	1.73	**1.492**	69486.601	18214.168	10070.666	4.154	1.763	**1.497**
10	50	0.8	85.032	25.567	8.414	6.154	2.173	**1.501**	44959.982	19429.209	12797.552	6.405	2.223	**1.509**	69527.885	18228.755	10074.035	6.501	2.242	**1.509**
1	50	0.9	22.801	6.951	**0.703**	2.657	1.366	1.393	271966.879	123807.12	8147.935	2.821	1.407	**1.406**	751440.542	418378.421	105601.218	2.949	1.444	**1.410**
5	50	0.9	110.602	32.634	3.429	8.52	2.783	**1.515**	272216.47	123954.815	8198.149	9.110	2.921	**1.513**	751857.779	418678.157	105852.306	9.589	3.018	**1.514**
10	50	0.9	223.867	65.53	6.668	16.46	4.793	**1.542**	272119.315	123804.205	8228.282	17.662	5.085	**1.530**	751521.878	418260.815	105688.226	18.564	5.238	**1.533**
1	50	0.95	33.631	10.103	**0.761**	3.376	1.532	1.436	259525.725	63099.131	1318.697	3.548	1.576	**1.435**	378257.454	62328.384	3624.173	3.587	1.582	**1.432**
5	50	0.95	163.61	47.601	3.455	13.223	4.054	**1.497**	259308.276	62984.916	1318.486	14.122	4.298	**1.489**	378000.4	62252.271	3634.782	14.358	4.322	**1.492**
10	50	0.95	319.828	90.852	6.899	22.63	6.208	**1.539**	259489.553	63068.59	1320.502	23.951	6.522	**1.534**	378210.213	62347.954	3646.722	24.527	6.674	**1.538**
1	100	0.8	8.016	2.682	**0.900**	1.211	1.023	1.138	14856.18	1785.804	1109.349	1.218	**1.024**	1.140	49141.245	9956.498	3139.145	1.224	**1.025**	1.143
5	100	0.8	40.411	12.628	4.317	2.308	**1.285**	1.416	14857.174	1790.82	1110.635	2.325	**1.287**	1.417	49114.256	9945.318	3137.823	2.349	**1.291**	1.420
10	100	0.8	80.741	25.239	8.529	3.739	1.686	**1.532**	14892.954	1805.585	1114.098	3.773	1.692	**1.533**	49165.728	9973.805	3144.359	3.811	1.697	**1.534**
1	100	0.9	14.385	4.531	**0.782**	1.458	1.090	1.238	49290.511	15515.865	2919.152	1.473	**1.093**	1.243	71044.994	16209.444	4777.114	1.496	**1.097**	1.250
5	100	0.9	69.687	20.61	3.67	3.355	1.536	**1.488**	49375.88	15551.268	2924.204	3.422	1.555	**1.491**	70991.087	16197.107	4780.66	3.531	1.578	**1.491**
10	100	0.9	140.746	41.436	7.265	5.571	2.050	**1.520**	49387.911	15562.731	2928.893	5.654	2.06	**1.519**	71333.569	16331.478	4792.289	5.898	2.127	**1.520**
1	100	0.95	27.765	8.287	**0.729**	1.942	1.199	1.364	97667.155	29058.228	1401.182	1.973	**1.207**	1.369	381563.018	188415.657	4475.914	2.011	**1.217**	1.375
5	100	0.95	135.335	38.539	3.433	5.633	2.085	**1.529**	97701.742	29054.883	1407.93	5.726	2.096	**1.529**	381755.21	188479.119	4477.368	5.862	2.124	**1.531**
10	100	0.95	279.99	82.078	6.700	10.147	3.199	**1.550**	97951.323	29170.418	1427.14	10.369	3.244	**1.553**	381791.619	188419.736	4479.598	10.69	3.322	**1.561**
1	200	0.8	8.202	2.668	**0.895**	1.118	1.018	1.092	1707.04	88.718	257.37	1.118	**1.018**	1.093	24028.221	7941.71	3735.021	1.120	**1.018**	1.094
5	200	0.8	42.141	13.247	4.313	1.728	**1.167**	1.336	1740.713	91.726	260.112	1.731	**1.167**	1.336	24039.073	7941.875	3734.765	1.743	**1.170**	1.338
10	200	0.8	84.549	26.201	8.524	2.425	**1.336**	1.461	1780.684	95.502	263.682	2.433	**1.339**	1.461	24058.517	7946.63	3736.432	2.447	**1.342**	1.460
1	200	0.9	14.696	4.537	**0.738**	1.217	1.035	1.151	22948.371	5544.276	956.247	1.219	**1.035**	1.152	44633.513	9553.022	1201.213	1.220	**1.034**	1.152
5	200	0.9	73.322	21.798	3.692	2.211	**1.276**	1.419	22950.016	5543.911	981.318	2.225	**1.278**	1.421	44615.021	9546.238	1201.764	2.261	**1.296**	1.423
10	200	0.9	147.872	43.559	7.195	3.704	1.660	**1.525**	23038.957	5582.191	1008.822	3.721	1.664	**1.526**	44693.284	9583.597	1205.384	3.728	1.663	**1.526**
1	200	0.95	31.562	9.17	**0.702**	1.517	1.110	1.276	35295.806	5531.456	638.058	1.521	**1.110**	1.276	124111.006	33503.082	1870.807	1.523	**1.110**	1.277
5	200	0.95	162.176	47.719	3.637	3.817	1.681	**1.528**	35410.954	5580.152	638.615	3.841	1.682	**1.528**	124218.044	33546.479	1863.449	3.868	1.689	**1.529**
10	200	0.95	314.547	89.211	6.875	6.052	2.125	**1.531**	35640.676	5671.806	643.211	6.078	2.129	**1.530**	124527.453	33700.271	1854.619	6.125	2.136	**1.528**

Table 3. The estimated MSE values of the considered estimators for the model when $p = 8$

			No outlier						One outlier						Two outliers					
σ^2	n	ρ	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**	**OLS**	**RE**	**LE**	**LSR**	**RRE**	**LRE**
1	50	0.8	21.089	5.793	1.509	3.755	**1.507**	1.835	139147.844	39737.4	9468.393	3.926	**1.539**	1.848	388397.223	139159.787	24145.655	4.155	**1.592**	1.864
5	50	0.8	106.443	28.954	7.558	14.335	3.739	**2.097**	139300.65	39785.071	9474.007	15.044	3.876	**2.093**	388623.289	139237.387	24149.224	15.918	4.042	**2.089**
10	50	0.8	208.961	55.83	15.185	27.613	6.508	**2.052**	139270.452	39762.681	9479.12	29.249	6.814	**2.040**	388525.391	139172.257	24155.227	30.819	7.109	**2.027**
1	50	0.9	35.800	9.638	**1.161**	6.039	2.017	2.013	242807.428	54992.036	4742.136	6.431	2.103	**2.021**	497556.099	121583.337	13511.857	6.555	2.126	**2.026**
5	50	0.9	183.061	49.255	5.776	23.738	5.628	**2.025**	242939.822	55058.251	4749.194	25.179	5.919	**2.012**	497559.784	121571.713	13513.03	26.199	6.149	**2.022**
10	50	0.9	360.535	95.785	11.446	46.145	10.312	**1.931**	242981.461	55032.333	4748.732	49.155	10.892	**1.920**	497814.968	121686.567	13524.012	50.782	11.185	**1.925**
1	50	0.95	102.571	27.969	**0.664**	15.941	4.263	2.131	479567.128	101864.307	2984.616	16.385	4.331	**2.119**	1142542.861	252811.176	5952.728	16.856	4.397	**2.111**
5	50	0.95	506.765	134.981	3.114	66.258	14.882	**1.861**	479603.915	101823.237	2985.231	69.592	15.601	**1.845**	1142701.658	252830.074	5953.948	72.775	16.400	**1.832**
10	50	0.95	1016.342	272.293	6.258	127.465	27.21	**1.704**	479642.683	101827.648	2988.019	132.843	28.202	**1.689**	1142611.687	252799.378	5957.66	138.349	29.28	**1.682**
1	100	0.8	21.952	6.325	1.459	2.345	**1.243**	1.665	35509.399	6622.481	3294.372	2.375	**1.248**	1.674	89907.325	19893.191	7589.355	2.413	**1.255**	1.687
5	100	0.8	112.031	31.04	7.279	7.852	2.422	**2.156**	35577.205	6640.059	3298.547	7.977	2.453	**2.160**	89975.62	19911.924	7593.955	8.195	2.505	**2.163**
10	100	0.8	218.515	59.876	14.193	14.36	3.682	**2.148**	35740.879	6695.676	3311.712	14.624	3.743	**2.146**	90108.298	19967.624	7608.083	15.140	3.869	**2.150**
1	100	0.9	41.298	11.274	**1.090**	3.62	1.504	1.911	56569.198	7152.557	1800.411	3.655	**1.509**	1.914	128857.612	20803.63	4113.459	3.706	**1.520**	1.921
5	100	0.9	203.455	54.211	5.328	13.419	3.55	**2.167**	56732.487	7193.956	1804.164	13.66	3.612	**2.167**	129042.961	20856.726	4118.128	13.878	3.645	**2.165**
10	100	0.9	400.159	104.894	10.413	26.425	6.337	**2.060**	56883.601	7234.105	1809.091	26.822	6.413	**2.055**	129052.864	20854.859	4119.762	27.291	6.512	**2.053**
1	100	0.95	76.929	20.342	**0.703**	6.198	2.035	2.106	126025.455	13035.134	727.896	6.276	**2.049**	2.107	190278.589	17611.911	1537.059	6.358	**2.064**	2.111
5	100	0.95	387.956	103.51	3.498	25.797	6.187	**2.092**	126070.222	13050.755	730.481	26.149	6.243	**2.084**	190356.293	17629.49	1537.948	26.723	6.389	**2.082**
10	100	0.95	799.35	213.639	7.16	49.689	10.981	**1.92**	126764.214	13228.797	743.915	50.439	11.154	**1.918**	191033.781	17772.868	1542.641	51.127	11.354	**1.916**
1	200	0.8	21.895	6.134	1.463	1.654	**1.111**	1.444	10012.152	843.329	870.751	1.656	**1.111**	1.445	28849.262	3309.752	2082.147	1.661	**1.112**	1.447
5	200	0.8	108.547	29.233	7.085	4.42	**1.683**	2.031	10084.569	857.11	875.245	4.440	**1.688**	2.033	28875.314	3316.304	2084.089	4.459	**1.690**	2.035
10	200	0.8	220.731	60.855	14.441	7.742	2.390	**2.203**	10186.622	877.742	881.913	7.773	2.394	**2.203**	29050.204	3361.629	2095.271	7.812	2.398	**2.201**
1	200	0.9	35.310	9.526	**1.098**	2.132	1.215	1.638	16948.57	1290.932	625.896	2.139	**1.217**	1.640	127540.396	38822.079	3355.094	2.150	**1.218**	1.643
5	200	0.9	180.901	48.881	5.537	6.55	**2.130**	2.148	17079.771	1316.061	630.099	6.584	**2.141**	2.149	127685.158	38866.985	3359.624	6.691	2.170	**2.153**
10	200	0.9	364.599	98.811	11.02	12.313	3.367	**2.225**	17254.997	1348.93	635.024	12.421	3.390	**2.226**	127877.033	38929.171	3365.18	12.616	3.420	**2.225**
1	200	0.95	80.118	21.452	**0.692**	3.538	1.510	1.952	95780.669	20585.028	676.354	3.572	**1.519**	1.958	220670.252	57282.291	1760.579	3.599	**1.524**	1.961
5	200	0.95	398.392	106.157	3.443	13.267	3.517	**2.191**	96094.469	20689.105	679.066	13.345	3.526	**2.188**	220860.583	57326.758	1762.045	13.542	3.557	**2.188**
10	200	0.95	806.223	215.746	6.852	25.588	6.081	**2.094**	96396.472	20776.776	681.414	25.801	6.122	**2.092**	221089.695	57358.344	1764.309	26.000	6.158	**2.090**

inclusion of the y-direction outliers, the performance of the commonly used OLS, RE and LE is quite poor. On the other hand, LSR, RRE and LRE exhibited different behaviors in different scenarios. The following observations can be obtained from Tables 1-3:

1. When the number of outliers is gradually increased along with the number of variables in the model by keeping ρ, n, and σ^2 constant, an increase in the estimated MSE values of all estimators is observed.
2. When n, p and σ^2 are held constant, we observe that the estimated MSE values of LSR generally increase as the correlation between variables is increased, while the estimated MSE values of RRE and LRE remain almost constant. On the other hand, when the correlation between the variables and the outliers in the data are increased, the estimated MSE values of the LSR and RRE estimators increase. On the other hand, for $p = 8$, the estimated MSE value of LRE decreases when the model variance is large.
3. When n, p and ρ are kept constant and the variance σ^2 is increased, the MSE values of the LSR, RRE and LRE estimators generally increase. When the model variance increases with the number of outliers, the estimated MSE values of the LSR and RRE increase. On the contrary, for $p = 8$, the estimated MSE values of LRE decrease at high correlation and small sample size.
4. When p, ρ and σ^2 are kept constant and the number of observations in the model is increased, a decrease is observed in the estimated MSE values of all estimators. When the number of outliers and the number of variables in the model are increased, a decrease is observed in the estimated MSE values of the LSR and RRE. On the other hand, the estimated MSE values of the LRE for $p = 8$ show an increase at high correlation and large variance values.

As a result, we can conclude that the estimated MSE values for LSR, RRE and LRE for variables such as n, p, ρ , and σ^2 with the change in the number of outliers are considerably lower than OLS, RE and LE.

In the second simulation scheme, we investigate the performance of LSR, RRE and LRE in the presence of y-direction outliers for each n, p, ρ, and σ^2. The purpose of this simulation is to investigate the performance of LSR, RRE and LRE with respect to MSE values given in (12) with various values of the biasing parameter k and d and the presence of outliers in the y-direction. The biasing parameters k and d are not estimated in the second simulation scheme. Only the MSE values obtained by increasing k and d values in the range $[0,\ 1]$ by 0.1 are compared. We only consider the cases $\rho = 0.8,\ 0.95$, $n = 50,\ \ 200$, and $p = 2,\ 8$, and $\sigma^2 = 1,\ \ 10$. Depending on these n, ρ, p, and σ^2 values, the explanatory variables are generated according to equation (10). Similar to the previous simulation scheme, we examine the effects of outliers in the y-direction on the estimators considering three different cases: no outliers, one outlier and two outliers. For every values of k and d, the simulation is run 2000 times. The results are collectively presented graphically in Figures 1-6.

Figures 1-6 clearly show the effects of varying the biasing parameter k and d between 0 and 1 on the estimated MSE values of the estimators. According to the figures, we can obtain the following results depending on each(n, ρ, p, σ^2).

1) The LSR estimator showed an increase in the estimated MSE values in the presence of none, one and two outliers in the y-direction, but generally showed a stable behavior.

2) Although the MSE values estimated for RRE decreased with increasing values of the biasing parameter k, it did not affect the MSE values estimated from the outliers in the y-direction.

3) Although the MSE values estimated for LRE increased with increasing values of the biasing parameter d, it did not affect the MSE values estimated from the outliers in the y-direction.

As a result, no dramatic change is observed in the MSE values estimated by comparing LSR, RRE and LSR among themselves as OLS, RE and LE. On the other hand, for large values of the biasing parameter k, RRE and for small values of the biasing parameter d, LRE stand out due to their performance.

4. AN EMPIRICAL APPLICATION

In this section, we created an experimental dataset to study the performance of LSR, RRE and LRE. To do this, we created a dataset using Equation (10) with $n = 100$,$p = 4$ and $\rho = 0.95$. We used set.seed(4) in the R Program. Using equation (11) to create the response variable with $\sigma^2 = 5$. Modified observations $y\,(1) = 500$ and $y\,(n) = 500$ were used to create two outliers. In this case, the eigenvalues of the $X'X$ matrix were calculated as 100.000, 3.738, 0.106, 0.092, and 0.064. The condition number is approximately 39.410, therefore the matrix X is moderate ill-conditioned. The eigenvalues of the $\underline{X}'\underline{X}$ matrix were calculated as 164.869, 5.437, 0.122, 0.077, and 0.060. The condition number is approximately 52.593, therefore the matrix $\underline{X}$ is moderate ill-conditioned. The numerical results are summarized in Table 4.

From Table 4, it can be observed that the estimated MSE values of LSR, RRE, and LRE give smaller values compared to OLS, RE, and LE. As a result, RRE and LRE outperform LSR in the presence of multicollinearity and y-direction outliers. It also seems that LRE can be a strong alternative to RRE.

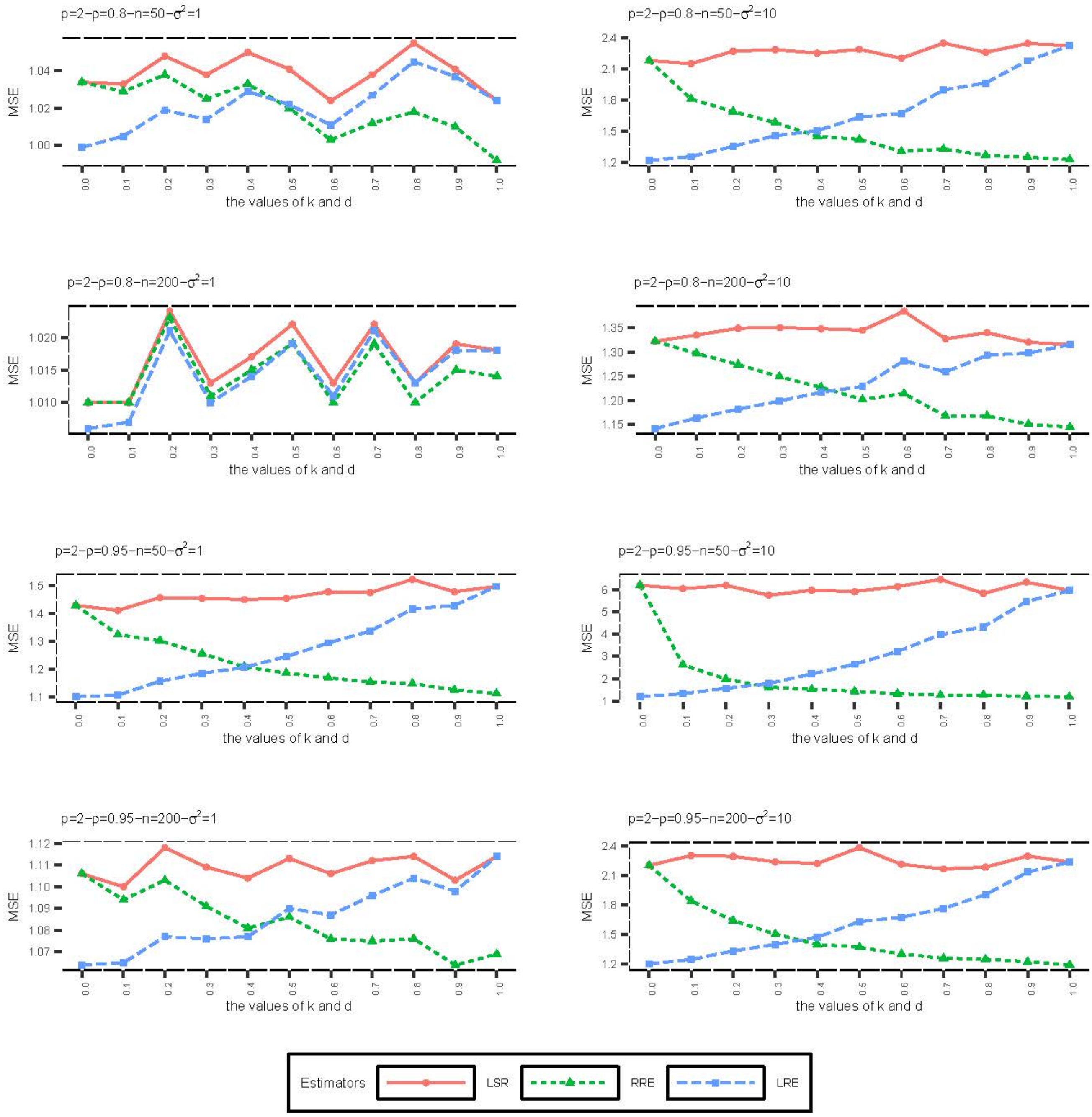

Figure 1. The estimated MSE values of LSR, RRE and LRE as a function k and d where $p = 2$ with no outlier

Table 4. The estimated parameter values and the estimated MSE values of the estimators

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$MSE(\hat{\beta})$
$\hat{\beta}_{OLS}$	11.0549	59.8429	455.3388	26.2779	-451.0035	83022.074
$\hat{\beta}_{RE}\left(\hat{k}_{RE} = 0.7087\right)$	10.9771	14.8855	67.4696	22.2895	-30.0934	1255.141
$\hat{\beta}_{LE}\left(\hat{d}_{LE} = 0\right)$	10.9454	14.598	53.2606	20.1941	-18.1441	777.143
$\hat{\beta}_{LSR}$	0.1297	6.8485	-4.443	-5.1036	1.8292	19.359
$\hat{\beta}_{RRE}\left(\hat{k}_{RRE} = 0.2234\right)$	0.1239	1.2467	-1.202	-1.4109	0.6017	1.224
$\hat{\beta}_{LRE}\left(\hat{d}_{LRE} = 0\right)$	0.1253	0.2183	-0.439	-0.5042	0.0808	0.253

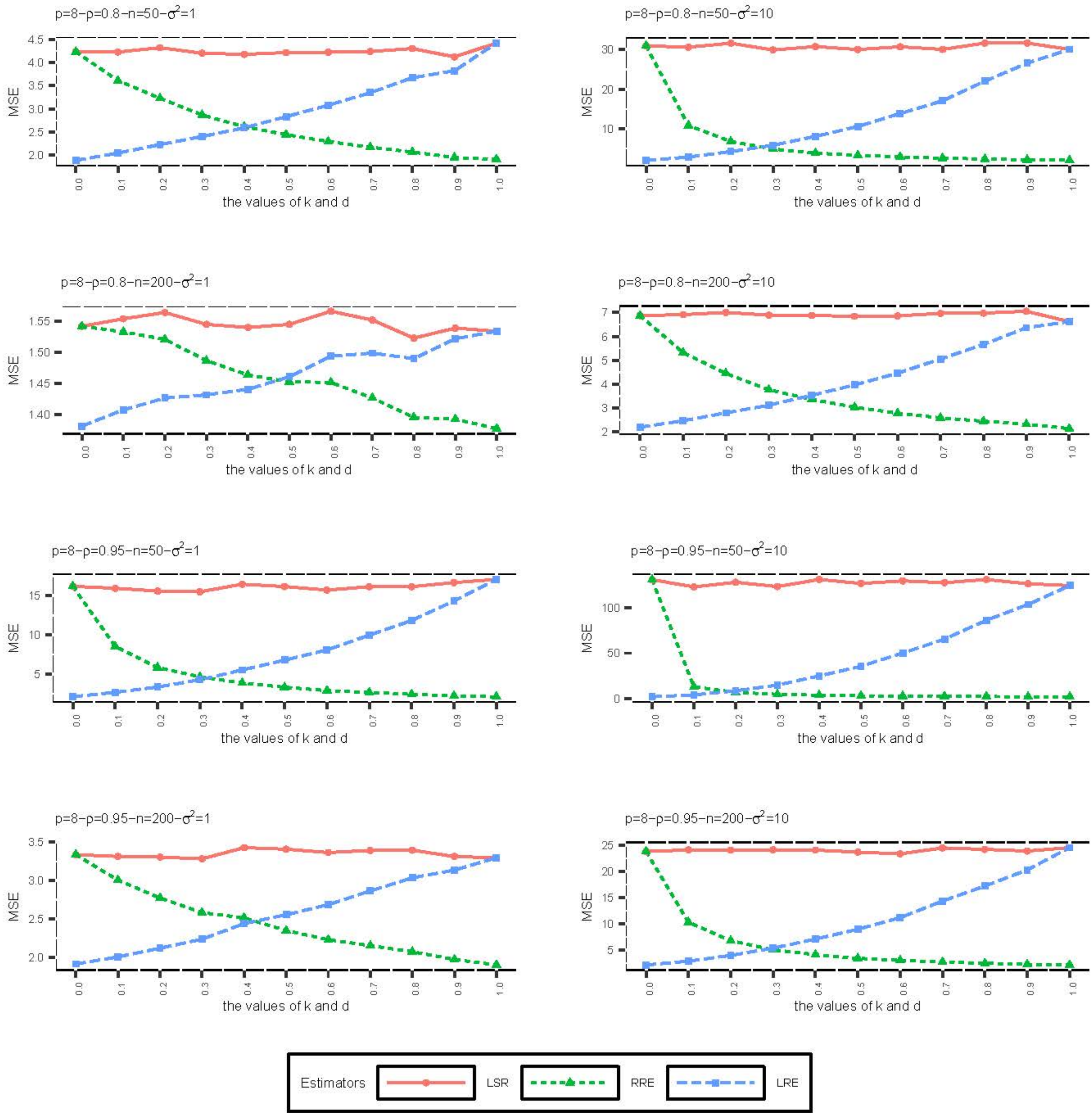

Figure 2. The estimated MSE values of LSR, RRE and LRE as a function k and d where $p = 8$ with no outlier

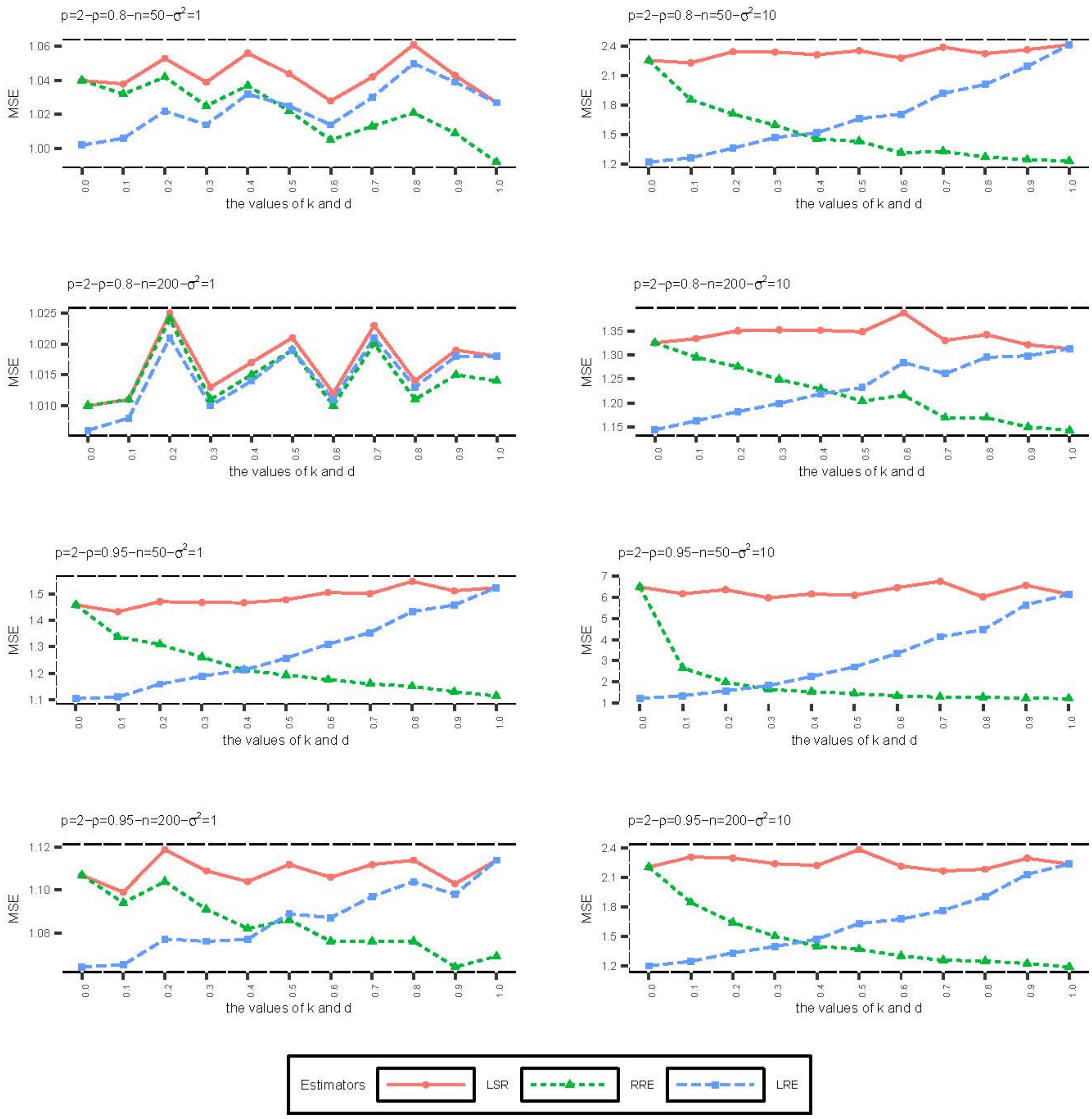

Figure 3. The estimated MSE values of LSR, RRE and LRE as a function k and d where $p = 2$ with one outlier

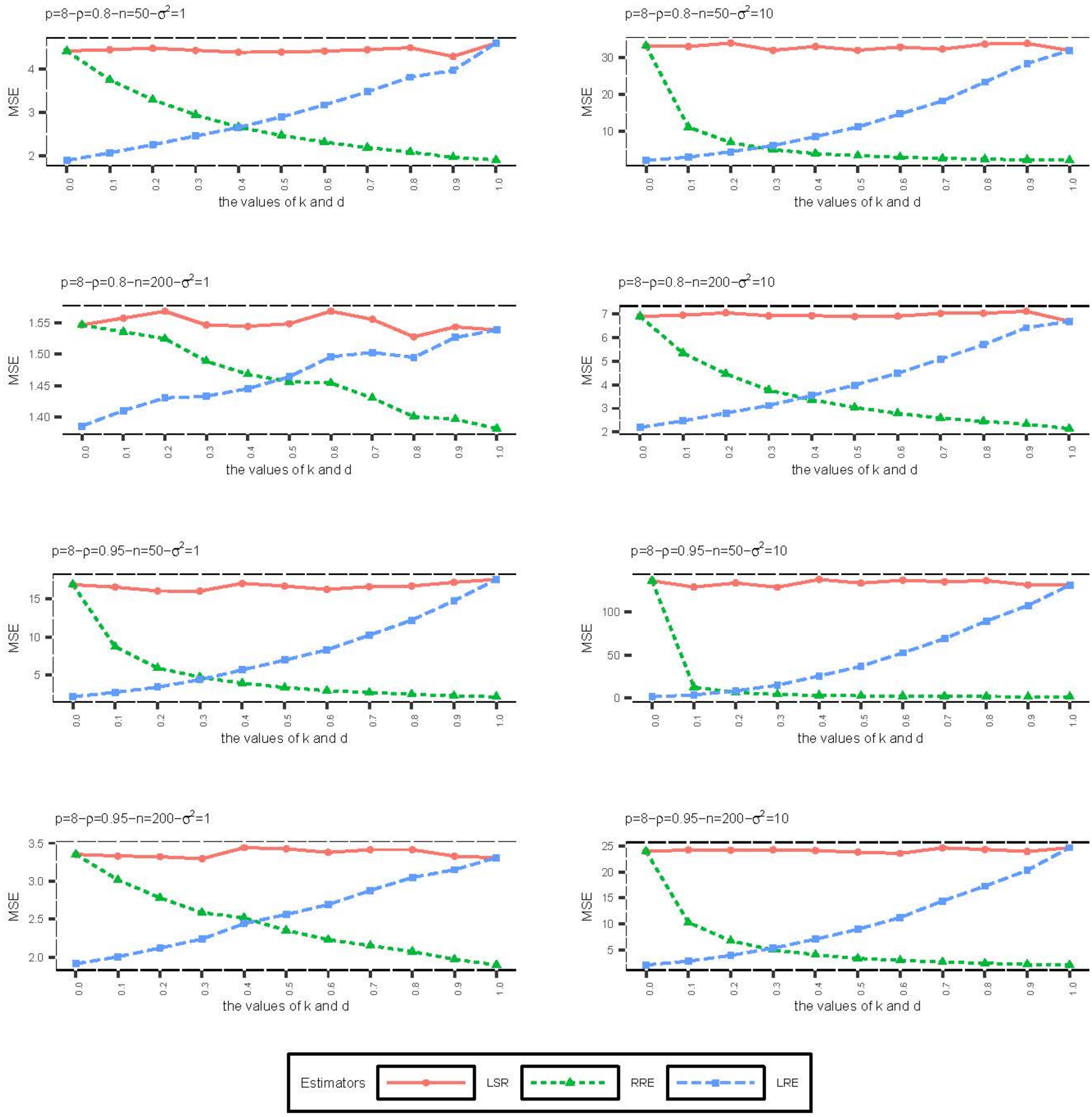

Figure 4. The estimated MSE values of LSR, RRE and LRE as a function k and d where $p = 8$ with one outlier

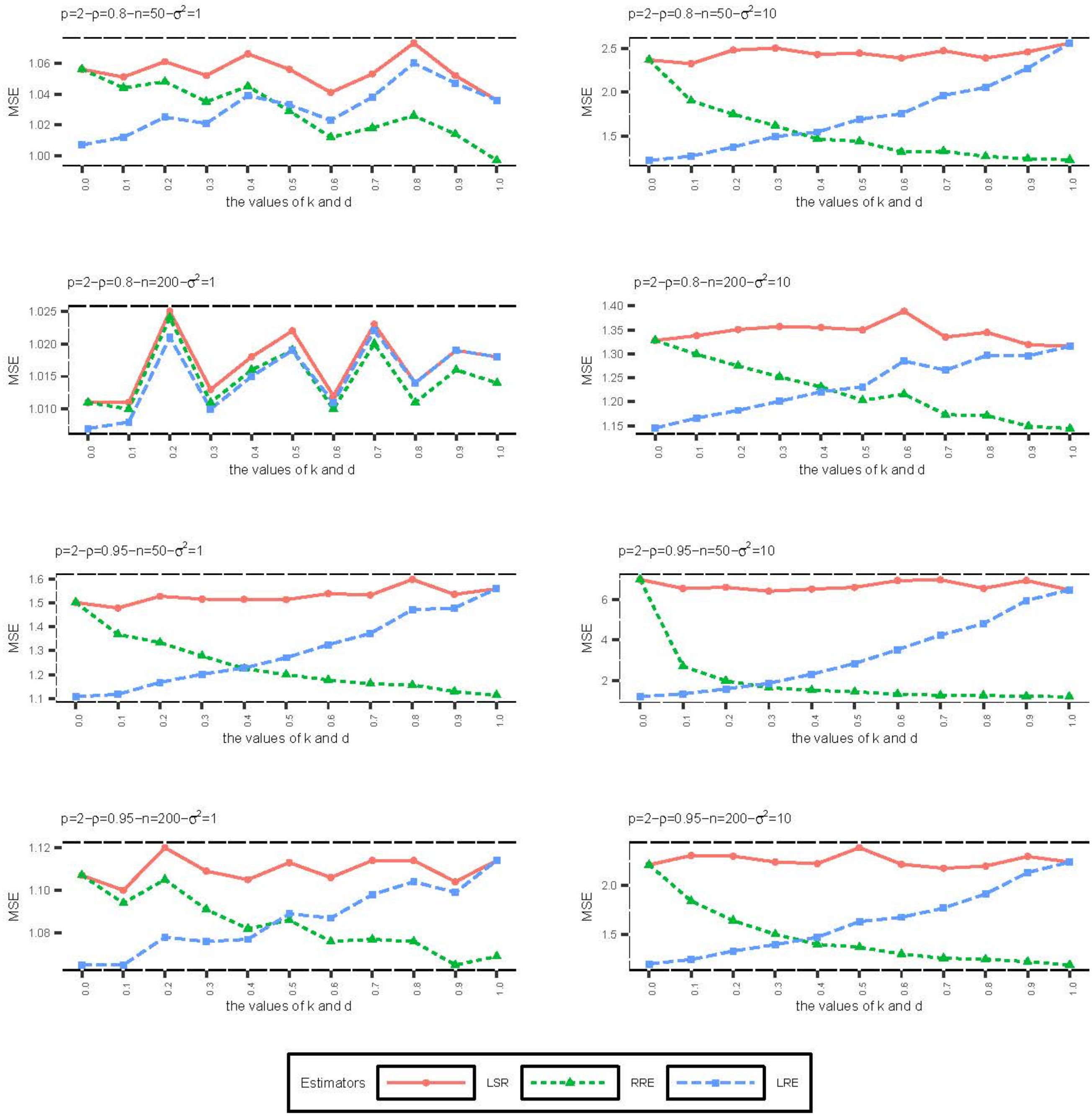

Figure 5. The estimated MSE values of LSR, RRE and LRE as a function k and d where $p = 2$ with two outliers

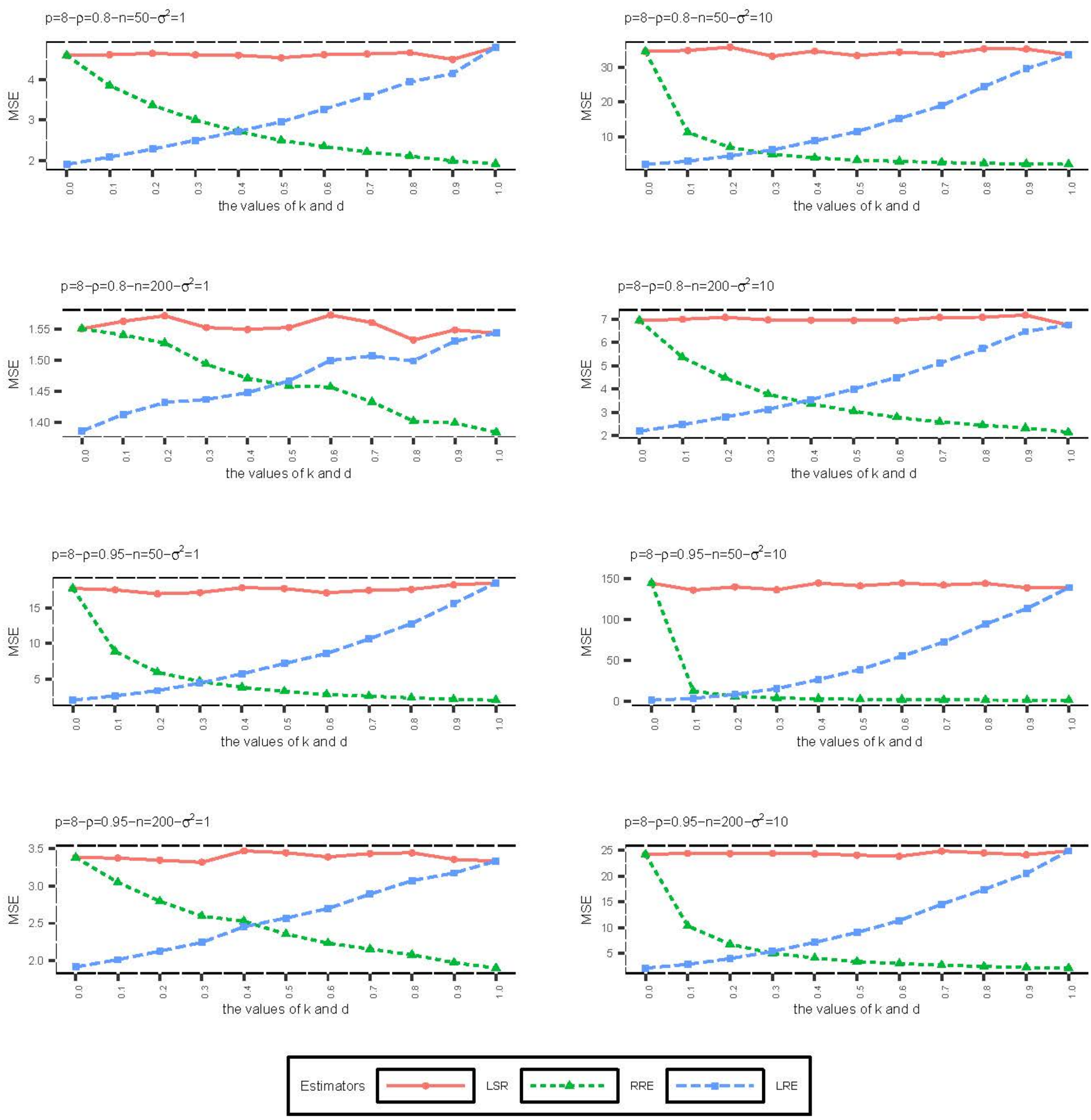

Figure 6. The estimated MSE values of LSR, RRE and LRE as a function k and d where p = 8 with two outliers

5. CONCLUSION

In this article, we proposed a new estimator named the LRE as an alternative to LSR and RRE in the presence of multicollinearity and y-direction outliers. Two separate Monte Carlo simulation study are conducted to examine the performance of LRE. In the first simulation study, we compared the considered estimators together with the estimates of the biasing parameters k and d. When the y-direction outliers are taken into account, the performance of OLS, RE and LE is considerably poor, while the performance of LSR, RRE and LRE is more stable. In the second simulation study, the performance of LSR, RRE and LRE are analyzed by choosing k and d values as fixed and equally spaced. According to the simulation results, LRE performs better for small values of d and RRE performs better for large values of k. According to the simulation results and the analysis of synthetic data, we recommend LRE as an alternative to RRE in the presence of y-direction outliers and multicollinearity between variables.

ACKNOWLEDGMENTS

The authors would like to thank the editor and reviewers for their valuable comments and suggestions that contributed to the improvement of the manuscript.

Peer Review: Externally peer-reviewed.
Conflict of Interest: Authors declared no conflict of interest.
Financial Disclosure: Authors declared no financial support.

LIST OF AUTHOR ORCIDS

İ. M. Giresunlu https://orcid.org/0009-0006-8737-1016
K. U. Akay https://orcid.org/0000-0002-8668-2879
E. Ertan https://orcid.org/0000-0002-6020-8749

REFERENCES

Akbilgic O., Akinci E. D., 2009, A Novel Regression Approach: Least Squares Ratio, Communications in Statistics-Theory and Methods, 38, 1539–1545.
Arslan, O., Billor, N., 2000, Robust Liu Estimator for Regression based on an M-estimator, Journal of Applied Statistics, 27, 1, 39–47.
Ertaş, H., Kaçıranlar, S., Güler, H., 2017, Robust Liu-type estimator for regression based on M-estimator, Communications in Statistics-Simulation and Computation, 46, 5, 3907–3932.
Filzmoser, P., Kurnaz, F. S., 2018, A robust Liu regression estimator, Communications in Statistics-Simulation and Computation, 47, 2, 432–443.
Hoerl A.E., Kennard R.W., 1970, Ridge regression: biased estimation for nonorthogonal problems, Technometrics, 12, 1, 55–67.
Jadhav, N. H., Kashid, D. N., 2016, Robust Linearized Ridge M-estimator for Linear Regression Model, Communication in Statistics–Simulation and Computation, 45, 3, 1001–1024.
Jadhav, N. H., Kashid, D. N., 2018, Ridge Least Squares Ratio Estimator For Linear Regression Model, International Journal of Agricultural & Statistical Sciences, 14, 2, 439–447.
Kan, B., Alpu, O., Yazici, B., 2013, Robust Ridge and Robust Liu Estimator for Regression based on the LTS Estimator, Journal of Applied Statistics, 40, 3, 644–655.
Kibria, B. G., 2003, Performance of some new ridge regression estimators, Communications in Statistics: Simulation and Computation, 32, 2, 419–435.
Liu, K., 1993, A new class of biased estimate in linear regression, Communications in Statistics-Theory and Methods, 22, 2, 393–402.
Maronna, R. A., Martin, R. D., Yohai, V. J., 2006, Robust Statistics Theory and Methods, Wiley, U.K.
Maronna, R.A., 2011, Robust ridge regression for high-dimensional data, Technometrics, 53, 1, 44–53.
McDonald G.C., Galarneau D.I., 1975, A Monte Carlo evaluation of some ridge-type estimators, Journal of the American Statistical Association, 70, 350, 407–416.
Montgomery, D. C., Peck, E. A., Vinig, G. G., 2001, Introduction to Linear Regression Analysis, 3th. Ed. John Wiley & Sons, Inc., USA.
Qasim, M., Amin, M., Omer, T., 2020). Performance of some new Liu parameters for the linear regression model, Communications in Statistics-Theory and Methods, 49, 17, 4178–4196.
Rousseeuw P.J., Leroy, A.M., 1987, Robust Regression and Outlier Detection, Wiley, New York.
Silvapulle, M. J., 1991, Robust Ridge Regression based on an M-estimator, Australian Journal of Statistics, 33, 3, 319–333.