A Simulation Study on Robust Alternatives of Least Squares Regression

Maw Maw Khin[1]

Abstract

Five methods of regression namely the ordinary least squares, least absolute value, *M*, least median squares and least trimmed squares are applied to the multiple regression model. The several distributional assumptions of errors are considered in this study. The required data sets are generated by using multiple linear regression models with three explanatory variables. Then, these data sets are transformed into outlier contaminated data sets. After that, the performances are compared in terms of bias and mean squared errors criteria and then the most suitable estimation method is chosen. Same sets of simulated data are used and mean squared errors and bias of these methods are compared. It is found that ordinary least squares estimation under a heavy-tailed distribution does not yield outlier robust estimates. Indeed, not only with the Gaussian distribution but also with the skewed distributions, ordinary least squares estimators collapse in the presence of small levels of outlier contamination. The Huber *M*-estimate and bisquare *M*-estimate estimate have shown to be more appropriate alternatives to the ordinary least squares in heavy-tailed distributions whereas the LMS estimates are better choices for skewed data. One best method could not be suggested in all situations; however the use of more than one method of exploratory data analysis is recommended in practice.

Keywords: Robust Estimators, Ordinary Least Squares, Heavy Tailed Distributions, Skewed Distribution, Gaussian Distribution.

1. Introduction

Modeling data by the means of linear least squares method is very important and crucial but the well-known ordinary least squares (OLS) regression procedure is only optimal under certain distributional assumption of errors. In practice, this assumption may not hold because of possibility of the skewness or presence of outliers in data. In theory, the assumption of normality does not meet, the standard least squares estimation for the regression coefficients $\beta s'$ will be biased and / or in-efficient [Hampel et al., (1986)] [2].

To overcome this problem, several alternative methods of the standard least squares regression (robust procedures) have been proposed. Among these, four methods *M*-estimation (based on Huber and Turkey weight function), Least Absolute Value estimation (LAV), Least Median Squares estimation (LMS) and Least Trimmed Squares estimation (LTS) methods are

[1] Professor and Head, Department of Statistics, Yangon University of Economics

used in this study. The aim of this paper is to make a comparison of these methods through a simulation study.

2. Data and Methods

In this study, the simulation data were used and transformed these data into the outliers contaminated data to make a comparison among traditional and robust estimation methods. These simulation data were generated based on the multiple regression model with three explanatory variables. The OLS and robust estimations methods were applied to these simulated data to estimate the parameters of the multiple regression model. For comparing the properties of estimation procedures, the mean squared errors (MSE) and bias criteria were used.

3. Results and Discussion

To analyze the effects of outliers on parameter estimation in regression, a simulation study was carried out. When performing such a simulation study, different error structures were taken into account in this study. The multiple linear regression model with three explanatory variables was applied. The multiple regression model is

$$Y = X\beta + e$$

where, $\boldsymbol{Y}$ is an (n×1) vector of observations with the design matrix $\boldsymbol{X}$ of the order n×p such that X_{i1} = 1, i = 1,…, n. $\boldsymbol{\beta}$ is a (p × 1) vector parameter and $\boldsymbol{e}$ is an (n×1) vector of errors. In this above model, the intercept and slopes were equal to one. These explanatory variables (p = 3) were generated from standard normal distribution. In this simulation study, the errors contained outliers were generated using heavy-tailed distribution (compare to standard normal distribution (*N*)) such as logistic (*LOG*), Cauchy(*C*) and skewed independent data sets like gamma (*GAM*) and exponential distribution (*EXP*). Thus, the errors were simulated from the following densities: *N* (0, 1), *LOG* (0, 1), *EXP* (1), *C* (0, 1), and *GAM* (1, 0.5). Table 1 shows the notations and parameters of distributions, which were used in the simulation process.

In each case, 10 replications were simulated and regression coefficients of OLS, LAV, Huber and Turkey *M*-estimates, LMS and LTS were calculated. To compare the properties of the estimation procedures, the mean squared errors (MSE) and bias of the estimated coefficients were computed using the following formulas

$$MSE = \frac{1}{10}\sum_{i=1}^{10}\left(\hat{\beta}_i - \beta\right)^2$$

$$Absolute\ \ Bias = \frac{1}{10}\sum_{i=1}^{10}\left|\hat{\beta}_i - \beta_i\right|$$

Table (1) Notations and Parameters of Distribution

Distribution	Notations and Parameters	p.d.f. [$f(x)$]
Normal	$x \sim N\left(\mu,\sigma^2\right),$ $-\infty < \mu < +\infty,\ \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2}, -\infty < x < +\infty$
Logistic	$x \sim LOG(\theta,\eta),\ \theta > 0$	$\frac{1}{\theta}\frac{\exp\left[(x-\eta)/\theta\right]}{\{1+\exp\left[(x-\eta)/\theta\right]\}^2}, -\infty < x < +\infty$
Exponential	$x \sim EXP(\theta), \theta > 0$	$\frac{1}{\theta}\exp\left(\frac{-x}{\theta}\right),\ x > 0$
Cauchy	$x \sim C(\mu,\sigma),$ $-\infty < \mu < +\infty, \sigma > 0$	$\frac{1}{\pi\sigma\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]},\ -\infty < x < +\infty$
Gamma	$x \sim GAM(\theta,k), \theta > 0,\ k > 0$	$\frac{1}{\theta^k\Gamma(k)} x^{k-1} e^{-x/\theta},\ x > 0$

Overall results of the methods under study and corresponding MSE and bias of 10 simulations for each estimation method were shown in Tables (2) to (6) and Figures (1) to (5). These figures illustrate the results of MSE and bias for the coefficients of multiple linear regression model with three explanatory variables ($p = 3$).

Based on the results of normal distribution, the bias of OLS is the smallest as expected, followed by the bias of Turkey and Huber-*M*, respectively. Moreover, in this case, the MSE of OLS is the smallest followed by the values of MSE of Huber and Turkey-*M*, respectively. Under the normal error distribution, it is found that the OLS method is more efficient than the robust methods. Thus, the low bias and MSE values of the OLS method are in line with the asymptotic robustness properties. In this normal distribution, the bias and MSE of LMS are much greater which followed by the biases and MSEs of LTS and LAV methods. The LMS method performs much worst in this case.

Regarding the logistic distribution, the bias of OLS, Turkey-*M* and Huber-*M* are close to each other and perform better than LAV, LTS and LMS methods. In this case, the MSE of OLS, Turkey and Huber-*M*, LAV and LTS methods are much close to each other but this value for LMS is significantly larger. Furthermore, although biases and MSEs of OLS, Turkey and Huber-*M* are significantly smaller than the bias and MSE of LMS, their patterns as shown in Figure 2(a) to (h) are intermingled and so no methods have a preferable bias and MSE in this situation. The bias and MSE of LMS are much larger than the others.

In exponential distribution, the LAV, LTS and Turkey-*M* are close to each other, but inferior to the Huber-*M* in terms of intercept. The bias of LMS is the smallest in this case. The OLS method as shown in part (a) of Figure 3 performs much worst in these situations. The MSE of OLS, for this situation is much greater which followed by the MSE values of Huber-*M*, Turkey-*M*, LTS and LAV respectively. The MSE of LMS is the smallest in this case. As indicated in Figure 3(c) to (h), the general pattern of the bias and MSE values for all methods are intermingled so that no preferred method could be chosen for the study of slope coefficients.

In Cauchy distribution, the biases of the robust methods for the intercepts are so close to each other, but the OLS method as shown in Figure 4(a) performs much worst in this situation. Furthermore, the MSE of OLS is significantly larger than the MSE of robust methods. The robust methods are so close to each other and their pattern as shown in Figure 4(b) are intermingled. So, no methods have a preferable MSE in this case. The similar results are found in the study of slope coefficients. The OLS method performs much worst based on bias criterion in this study. The biases of LAV, LTS, Turkey and Huber-*M* and LMS are so close to each other. In this case, the MSE of OLS is significantly larger than the MSE of robust methods. From this study, it is found that, the general patterns of the bias and MSE values for all robust methods are intermingled so that no preferred method can be selected for this case.

Concerning the case of gamma distribution, the intercept of LAV, LTS and Turkey-*M* are close to each other, but lower to the Huber-*M* depending on bias criterion. In this situation, the bias of LMS is significantly smaller than the bias of other methods. The OLS method as described in Figure 5(a) performs much worst. In addition, in this case, the MSE of LAV, LTS and Turkey-*M* are close to each other. The MSE of LMS is the smallest and performs better than the other methods. The OLS method as described in Figure 5(b) performs much poorest.

Moreover, the bias and MSE of Turkey-*M* are the smallest in terms of the slope. It is closely followed by the bias and MSE values of Huber-*M*, LAV, LTS and OLS.

Table (2) Performances of OLS and Robust Methods of Normal Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =10	OLS	Bias	0.3086	0.2048	0.2738	0.2051
		MSE	0.1757	0.0644	0.1107	0.1176
	LAV	Bias	0.4002	0.2181	0.3676	0.2434
		MSE	0.2544	0.0695	0.2465	0.0907
	M-Huber	Bias	0.2974	0.2060	0.2266	0.2026
		MSE	0.1453	0.0685	0.0938	0.1038
	M-Turkey	Bias	0.3015	0.3553	0.3434	0.3000
		MSE	0.1488	0.3310	0.2575	0.1934
	LTS	Bias	0.3432	0.2608	0.3072	0.2854
		MSE	0.1584	0.1668	0.1986	0.1045
	LMS	Bias	0.5234	0.5989	0.5645	0.4626
		MSE	0.4554	0.5986	0.5141	0.8200
n =20	OLS	Bias	0.1912	0.1865	0.2020	0.1403
		MSE	0.0602	0.0617	0.0744	0.0276
	LAV	Bias	0.2116	0.2230	0.2501	0.2515
		MSE	0.0602	0.0777	0.1012	0.1050
	M-Huber	Bias	0.2020	0.2012	0.2022	0.1145
		MSE	0.0701	0.0647	0.0720	0.0226
	M-Turkey	Bias	0.1963	0.1941	0.2146	0.1186
		MSE	0.0641	0.0634	0.0719	0.0235
	LTS	Bias	0.2531	0.2828	0.2812	0.2082
		MSE	0.1131	0.1352	0.1375	0.0776
	LMS	Bias	0.5271	0.4909	0.4668	0.3154
		MSE	0.4082	0.3397	0.3225	0.1779
n = 30	OLS	Bias	0.1642	0.1703	0.1450	0.1526
		MSE	0.0376	0.0494	0.0304	0.0334
	LAV	Bias	0.2056	0.2868	0.1536	0.1546
		MSE	0.0603	0.1480	0.0319	0.0306
	M-Huber	Bias	0.1940	0.1567	0.1554	0.1365
		MSE	0.0473	0.0446	0.0286	0.0249
	M-Turkey	Bias	0.1860	0.1578	0.1491	0.1342
		MSE	0.0443	0.0441	0.0273	0.0255
	LTS	Bias	0.2227	0.1474	0.1598	0.0857
		MSE	0.0677	0.0400	0.0319	0.0132
	LMS	Bias	0.2395	0.1996	0.3296	0.3494
		MSE	0.0948	0.0549	0.1444	0.2118
n =50	OLS	Bias	0.1609	0.2201	0.1345	0.0908
		MSE	0.0328	0.0588	0.0265	0.0104
	LAV	Bias	0.1434	0.2307	0.1200	0.0979
		MSE	0.0284	0.0807	0.0200	0.0132
	M-Huber	Bias	0.1602	0.2016	0.1367	0.0905
		MSE	0.0315	0.0586	0.0254	0.0097
	M-Turkey	Bias	0.1643	0.1964	0.1331	0.0857
		MSE	0.0331	0.0550	0.0242	0.0090
	LTS	Bias	0.1679	0.2641	0.1527	0.0923
		MSE	0.0388	0.0878	0.0316	0.0106
	LMS	Bias	0.2383	0.3356	0.1687	0.2540
		MSE	0.0977	0.1592	0.0445	0.1204
n =80	OLS	Bias	0.1396	0.1424	0.0917	0.0710
		MSE	0.0232	0.0279	0.0110	0.0108
	LAV	Bias	0.1446	0.1501	0.1034	0.0653
		MSE	0.0266	0.0358	0.0161	0.0075
	M-Huber	Bias	0.1358	0.1374	0.0877	0.0641
		MSE	0.0231	0.0294	0.0106	0.0098
	M-Turkey	Bias	0.1413	0.1304	0.0893	0.0611
		MSE	0.0252	0.0273	0.0108	0.0010
	LTS	Bias	0.1184	0.1761	0.1093	0.1218
		MSE	0.0207	0.0478	0.0168	0.0268
	LMS	Bias	0.2239	0.1566	0.1329	0.2327

		MSE	0.0748	0.0343	0.0397	0.0876
n =100	OLS	Bias	0.1253	0.1307	0.0826	0.0697
		MSE	0.0229	0.0217	0.0089	0.0092
	LAV	Bias	0.1559	0.1067	0.0858	0.0728
		MSE	0.0395	0.0184	0.0108	0.0080
	M-Huber	Bias	0.1308	0.1307	0.0742	0.0696
		MSE	0.0255	0.0230	0.0080	0.0090
	M-Turkey	Bias	0.1330	0.1255	0.0777	0.0705
		MSE	0.0266	0.0217	0.0086	0.0091
	LTS	Bias	0.1249	0.1651	0.0681	0.0894
		MSE	0.0290	0.0333	0.0070	0.0102
	LMS	Bias	0.2392	0.2121	0.2529	0.1743
		MSE	0.0761	0.0825	0.0919	0.0394

Source: Calculations based on simulated data

Table (3) Performances of OLS and Robust Methods of Logistic Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n = 10	OLS	Bias	0.6008	0.5992	0.6534	0.6685
		MSE	0.5172	0.6685	0.6845	0.7979
	LAV	Bias	0.6979	0.7139	0.8887	0.8093
		MSE	0.8873	0.9214	1.1780	1.0937
	M-Huber	Bias	0.6727	0.5797	0.7264	0.7172
		MSE	0.8082	0.7134	0.8194	0.9156
	M-Turkey	Bias	0.7320	0.6707	0.7471	0.8297
		MSE	0.8665	1.0874	0.8327	1.3582
	LTS	Bias	0.8595	0.5997	0.9042	0.9234
		MSE	1.1525	0.7990	1.1726	1.2751
	LMS	Bias	1.3391	0.7972	1.1474	1.3129
		MSE	2.3115	1.1169	1.4996	2.7096
n = 20	OLS	Bias	0.2985	0.4272	0.4600	0.3331
		MSE	0.1225	0.2685	0.2970	0.1586
	LAV	Bias	0.3014	0.4991	0.4726	0.4228
		MSE	0.1230	0.4735	0.4369	0.2700
	M-Huber	Bias	0.2786	0.4063	0.4641	0.3586
		MSE	0.1150	0.2573	0.3216	0.1796
	M-Turkey	Bias	0.2655	0.4036	0.4596	0.3478
		MSE	0.1071	0.2581	0.3210	0.1744
	LTS	Bias	0.3326	0.5109	0.6293	0.4290
		MSE	0.1207	0.5581	0.5929	0.2601
	LMS	Bias	0.4507	0.6996	1.0963	0.9802
		MSE	0.2784	0.8511	1.6086	1.6227
n =30	OLS	Bias	0.2664	0.2581	0.3643	0.1759
		MSE	0.1078	0.2136	0.1852	0.0522
	LAV	Bias	0.2543	0.3313	0.3976	0.2949
		MSE	0.0914	0.2682	0.1793	0.1186
	M-Huber	Bias	0.2501	0.2630	0.3493	0.1980
		MSE	0.0871	0.2155	0.1685	0.0577
	M-Turkey	Bias	0.2459	0.2625	0.3455	0.1945
		MSE	0.0787	0.2153	0.1732	0.0555
	LTS	Bias	0.2799	0.3272	0.4033	0.4543
		MSE	0.1065	0.1966	0.1945	0.2414
	LMS	Bias	0.5637	0.9437	0.5797	0.3814
		MSE	0.3630	1.1606	0.4564	0.2530
n =50	OLS	Bias	0.1597	0.2760	0.2470	0.3024
		MSE	0.0493	0.1430	0.0891	0.1107
	LAV	Bias	0.2593	0.3333	0.2142	0.2767
		MSE	0.0869	0.1754	0.0626	0.1248
	M-Huber	Bias	0.1625	0.2748	0.2582	0.2643
		MSE	0.0448	0.1357	0.1042	0.0886
	M-Turkey	Bias	0.1554	0.2684	0.2747	0.2420
		MSE	0.0403	0.1312	0.1154	0.0813
	LTS	Bias	0.2061	0.2707	0.2730	0.2834
		MSE	0.0819	0.1287	0.1338	0.1440
	LMS	Bias	0.5312	0.5433	0.6406	0.4673
		MSE	0.3731	0.5662	0.8952	0.3356
n =80	OLS	Bias	0.1427	0.1660	0.2487	0.1713

		MSE	0.0305	0.0381	0.0849	0.0392
	LAV	Bias	0.1856	0.1278	0.2423	0.1935
		MSE	0.0646	0.0230	0.1177	0.0472
	M-Huber	Bias	0.1496	0.1269	0.2279	0.1816
		MSE	0.0329	0.0248	0.0833	0.0421
	M-Turkey	Bias	0.1524	0.1208	0.2276	0.1701
		MSE	0.0328	0.0219	0.0862	0.0384
	LTS	Bias	0.1775	0.1307	0.2478	0.1787
		MSE	0.0520	0.0251	0.1040	0.0510
	LMS	Bias	0.3531	0.2450	0.2971	0.3855
		MSE	0.1638	0.0851	0.1457	0.1698
n =100	OLS	Bias	0.1248	0.1285	0.1792	0.1459
		MSE	0.0185	0.0285	0.0525	0.0304
	LAV	Bias	0.1408	0.0968	0.1893	0.2056
		MSE	0.0289	0.0160	0.0593	0.0569
	M-Huber	Bias	0.1349	0.0985	0.1723	0.1672
		MSE	0.0228	0.0192	0.0559	0.0364
	M-Turkey	Bias	0.1412	0.0949	0.1700	0.1670
		MSE	0.0249	0.0163	0.0568	0.0373
	LTS	Bias	0.1737	0.1063	0.2237	0.2346
		MSE	0.0422	0.0196	0.0996	0.0734
	LMS	Bias	0.3532	0.2339	0.2436	0.3622
		MSE	0.1769	0.0805	0.1291	0.1676

Source: Calculations based on simulated data

Table (4) Performances of OLS and Robust Methods of Exponential Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =10	OLS	Bias	0.8869	0.4223	0.1918	0.2208
		MSE	0.8400	0.3801	0.0503	0.0698
	LAV	Bias	0.8068	0.4030	0.1430	0.1966
		MSE	0.7369	0.3578	0.0402	0.0632
	M-Huber	Bias	0.7993	0.4430	0.1904	0.2118
		MSE	0.6795	0.4550	0.0454	0.0675
	M-Turkey	Bias	0.7601	0.4563	0.2263	0.2151
		MSE	0.6432	0.4974	0.0705	0.0734
	LTS	Bias	0.6763	0.5282	0.2992	0.2817
		MSE	0.5482	0.7216	0.1790	0.2005
	LMS	Bias	0.7246	0.7137	0.3599	0.4490
		MSE	0.6232	0.8143	0.1961	0.3464
n =20	OLS	Bias	0.9875	0.1521	0.2115	0.1170
		MSE	1.0145	0.0323	0.0713	0.0157
	LAV	Bias	0.7316	0.1719	0.1408	0.1462
		MSE	0.6027	0.0533	0.0362	0.0288
	M-Huber	Bias	0.8483	0.1322	0.1651	0.1312
		MSE	0.7599	0.0231	0.0450	0.0207
	M-Turkey	Bias	0.8008	0.1512	0.1453	0.1400
		MSE	0.6883	0.0425	0.0403	0.0273
	LTS	Bias	0.7500	0.2065	0.1719	0.1410
		MSE	0.6225	0.0678	0.0560	0.0298
	LMS	Bias	0.4050	0.1926	0.2943	0.2452
		MSE	0.1790	0.0509	0.1541	0.1608
n =30	OLS	Bias	0.9582	0.1562	0.1160	0.1363
		MSE	0.9524	0.0279	0.0216	0.0311
	LAV	Bias	0.7201	0.1580	0.1097	0.1080
		MSE	0.5627	0.0354	0.0169	0.0189
	M-Huber	Bias	0.8055	0.1222	0.1107	0.1184
		MSE	0.7929	0.0246	0.0208	0.0207
	M-Turkey	Bias	0.7354	0.1253	0.1247	0.0962
		MSE	0.5590	0.0254	0.0232	0.0136
	LTS	Bias	0.7095	0.1063	0.0962	0.1040
		MSE	0.5175	0.0238	0.0126	0.0150
	LMS	Bias	0.4909	0.1890	0.1376	0.1206
		MSE	0.2991	0.0686	0.0308	0.0281
n =50	OLS	Bias	0.9627	0.1069	0.1066	0.1008
		MSE	0.9419	0.0137	0.0165	0.0132

	LAV	Bias	0.6673	0.0875	0.1037	0.0739
		MSE	0.4551	0.0127	0.0142	0.0090
	M-Huber	Bias	0.7972	0.0706	0.0975	0.0903
		MSE	0.6482	0.0073	0.0123	0.0099
	M-Turkey	Bias	0.6945	0.0716	0.1055	0.0732
		MSE	0.4922	0.0081	0.0143	0.0077
	LTS	Bias	0.6937	0.0684	0.1143	0.1075
		MSE	0.4889	0.0068	0.0170	0.0161
	LMS	Bias	0.4415	0.0908	0.0847	0.1069
		MSE	0.2040	0.0117	0.0140	0.0188
n =80	OLS	Bias	0.9860	0.0661	0.0961	0.0642
		MSE	0.9849	0.0060	0.0132	0.0046
	LAV	Bias	0.6673	0.0854	0.0918	0.0806
		MSE	0.4537	0.0134	0.0118	0.0091
	M-Huber	Bias	0.8195	0.0667	0.0769	0.0513
		MSE	0.6854	0.0063	0.0097	0.0030
	M-Turkey	Bias	0.6996	0.0743	0.0823	0.0474
		MSE	0.4996	0.0083	0.0093	0.0037
	LTS	Bias	0.7229	0.0657	0.0987	0.0494
		MSE	0.5319	0.0064	0.0175	0.0035
	LMS	Bias	0.4021	0.0549	0.0550	0.0742
		MSE	0.1784	0.0052	0.0055	0.0101
n =100	OLS	Bias	0.9979	0.0574	0.0942	0.0614
		MSE	1.0088	0.0052	0.0132	0.0061
	LAV	Bias	0.6812	0.0877	0.1129	0.0625
		MSE	0.4737	0.0142	0.0162	0.0055
	M-Huber	Bias	0.8355	0.0667	0.0767	0.0415
		MSE	0.7119	0.0058	0.0097	0.0022
	M-Turkey	Bias	0.7175	0.0704	0.0853	0.0385
		MSE	0.5241	0.0080	0.0102	0.0025
	LTS	Bias	0.7280	0.0683	0.0822	0.0420
		MSE	0.5378	0.0066	0.0118	0.0026
	LMS	Bias	0.4288	0.0730	0.1042	0.0431
		MSE	0.1981	0.0077	0.0156	0.0032

Source: Calculations based on simulated data

Table (5) Performances of OLS and Robust Methods of Cauchy Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n = 10	OLS	Bias	2.6781	1.7431	2.8656	2.4600
		MSE	22.0394	8.3795	19.2347	17.3125
	LAV	Bias	0.9250	1.1730	1.2433	1.2245
		MSE	1.3160	2.7178	3.2583	2.4527
	M-Huber	Bias	1.1229	1.0392	1.7524	1.2515
		MSE	2.0275	1.6833	9.6928	2.2957
	M-Turkey	Bias	1.0268	0.7001	1.6003	1.0101
		MSE	1.7355	0.8209	10.472	1.5833
	LTS	Bias	0.9644	0.7875	1.2490	1.2296
		MSE	1.4443	1.0583	3.2550	2.6758
	LMS	Bias	0.7397	0.8135	0.9280	1.2161
		MSE	1.0858	1.2536	1.5602	3.3787
n = 20	OLS	Bias	2.2200	1.7870	1.6756	1.2306
		MSE	8.3993	6.6070	3.9963	3.4722
	LAV	Bias	0.3561	0.2274	0.5412	0.3553
		MSE	0.2330	0.0982	0.4833	0.1745
	M-Huber	Bias	0.3121	0.2656	0.5713	0.3249
		MSE	0.2045	0.1198	0.6117	0.1842
	M-Turkey	Bias	0.3683	0.3388	0.5146	0.3765
		MSE	0.1670	0.1502	0.4428	0.1782
	LTS	Bias	0.3011	0.3056	0.5078	0.2908
		MSE	0.1566	0.1157	0.4263	0.1383
	LMS	Bias	0.3995	0.2609	0.4618	0.4490
		MSE	0.2390	0.1140	0.3158	0.2748
n =30	OLS	Bias	4.3323	5.8791	1.5400	1.5029
		MSE	97.9254	235.9916	5.4072	9.3752
	LAV	Bias	0.2233	0.2265	0.4099	0.1742
		MSE	0.0657	0.0824	0.2798	0.0398

	M-Huber	Bias	0.2551	0.1866	0.5672	0.2752
		MSE	0.0884	0.0459	0.4956	0.1044
	M-Turkey	Bias	0.1766	0.1929	0.5003	0.3177
		MSE	0.0427	0.0554	0.3877	0.1151
	LTS	Bias	0.2158	0.2688	0.6248	0.3137
		MSE	0.0806	0.1131	0.6523	0.1682
	LMS	Bias	0.3576	0.3978	0.5022	0.4450
		MSE	0.2412	0.2607	0.3258	0.3863
n =50	OLS	Bias	2.4380	3.3922	1.2825	2.1126
		MSE	23.9028	67.9288	4.1653	20.6503
	LAV	Bias	0.1430	0.1777	0.2044	0.2255
		MSE	0.0331	0.0488	0.0637	0.0804
	M-Huber	Bias	0.1170	0.1531	0.2824	0.2241
		MSE	0.0227	0.0377	0.1051	0.0967
	M-Turkey	Bias	0.1387	0.1916	0.2494	0.2393
		MSE	0.0314	0.0454	0.0846	0.0848
	LTS	Bias	0.1595	0.2127	0.3489	0.2284
		MSE	0.0418	0.0522	0.1596	0.0937
	LMS	Bias	0.2338	0.2931	0.2854	0.2233
		MSE	0.0738	0.1162	0.1304	0.1053
n =80	OLS	Bias	1.5801	2.3042	1.0224	1.4368
		MSE	8.4181	31.4442	3.7895	7.2829
	LAV	Bias	0.1200	0.1266	0.0915	0.1574
		MSE	0.0197	0.0213	0.0193	0.0388
	M-Huber	Bias	0.1139	0.1350	0.1586	0.1872
		MSE	0.0231	0.0249	0.0342	0.0622
	M-Turkey	Bias	0.1237	0.2073	0.1480	0.2121
		MSE	0.0361	0.0481	0.0401	0.0670
	LTS	Bias	0.1428	0.2518	0.1406	0.2555
		MSE	0.0561	0.0658	0.0374	0.0884
	LMS	Bias	0.3140	0.3346	0.1615	0.1461
		MSE	0.1215	0.1532	0.0465	0.0392
n =100	OLS	Bias	1.5374	2.0641	1.0080	1.0555
		MSE	6.4297	22.3868	2.9185	3.0349
	LAV	Bias	0.1039	0.1206	0.0469	0.1557
		MSE	0.0154	0.0180	0.0073	0.0350
	M-Huber	Bias	0.0824	0.1186	0.1001	0.1559
		MSE	0.0112	0.0188	0.0162	0.0421
	M-Turkey	Bias	0.1182	0.1855	0.1105	0.1698
		MSE	0.0206	0.0421	0.0205	0.0457
	LTS	Bias	0.1355	0.2186	0.1286	0.1704
		MSE	0.0280	0.0627	0.0240	0.0406
	LMS	Bias	0.2323	0.1966	0.1657	0.2029
		MSE	0.0838	0.0522	0.0391	0.0536

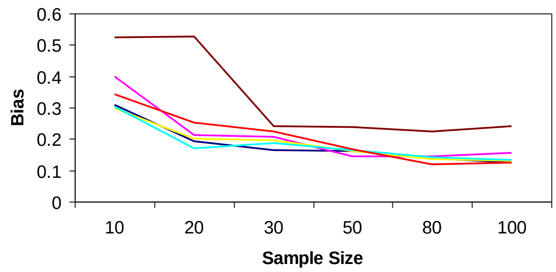
Source: Calculations based on simulated data

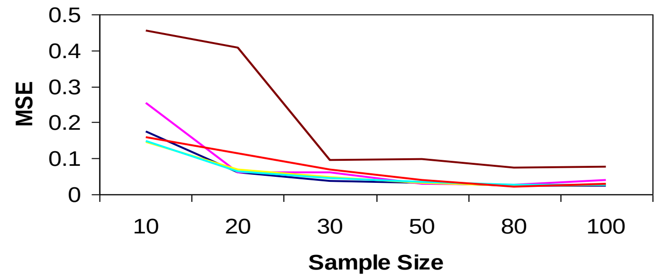
Table (6) Performances of OLS and Robust Methods of Gamma Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =10	OLS	Bias	2.3935	0.8216	0.6492	0.5599
		MSE	7.3810	1.2113	0.7848	0.4866
	LAV	Bias	2.3062	1.1021	0.6106	0.7862
		MSE	7.3690	2.8982	0.5796	1.0408
	M-Huber	Bias	2.2569	0.8284	0.6746	0.5358
		MSE	6.8373	1.2261	0.8309	0.4439
	M-Turkey	Bias	2.0283	0.9482	0.6500	0.5055
		MSE	5.6560	2.1358	0.7479	0.3862
	LTS	Bias	2.1743	1.0324	0.9422	0.6483
		MSE	6.5913	2.3942	2.3276	0.5631
	LMS	Bias	1.5070	0.8757	0.6243	0.4788
		MSE	2.7178	1.8015	0.4890	0.3060
n =20	OLS	Bias	2.1462	0.3684	0.3527	0.5616
		MSE	4.9348	0.2569	0.2105	0.4424
	LAV	Bias	1.5007	0.4235	0.3415	0.5842
		MSE	2.4972	0.2949	0.2167	0.4840
	M-Huber	Bias	1.9125	0.3363	0.3164	0.4549

		MSE	4.0267	0.2019	0.1985	0.2724
	M-Turkey	Bias	1.7407	0.3463	0.3121	0.4594
		MSE	3.5293	0.2229	0.1919	0.2590
	LTS	Bias	1.8206	0.4639	0.4945	0.5835
		MSE	3.8613	0.3158	0.6900	0.4384
	LMS	Bias	1.0030	0.3097	0.4122	0.5455
		MSE	1.3840	0.1284	0.3720	0.5856
n =30	OLS	Bias	2.0303	0.3283	0.1996	0.4099
		MSE	4.2804	0.1445	0.0567	0.3006
	LAV	Bias	1.4922	0.2176	0.1931	0.3677
		MSE	2.4788	0.0725	0.0701	0.2350
	M-Huber	Bias	1.7707	0.2070	0.1685	0.3525
		MSE	3.2870	0.0587	0.0403	0.1965
	M-Turkey	Bias	1.5254	0.1952	0.1656	0.2685
		MSE	2.5688	0.0571	0.0462	0.1467
	LTS	Bias	1.5158	0.2370	0.2220	0.2505
		MSE	2.3865	0.0768	0.0645	0.1009
	LMS	Bias	1.1215	0.3789	0.3694	0.5825
		MSE	1.7866	0.3366	0.2039	0.5754
n=50	OLS	Bias	1.9751	0.2472	0.1459	0.2949
		MSE	4.0195	0.0927	0.0295	0.1209
	LAV	Bias	1.4359	0.1300	0.1245	0.3035
		MSE	2.2758	0.0249	0.0304	0.1276
	M-Huber	Bias	1.6864	0.1899	0.0695	0.2437
		MSE	2.9755	0.0484	0.0103	0.0785
	M-Turkey	Bias	1.4592	0.1921	0.0731	0.2034
		MSE	2.3139	0.0459	0.0086	0.0483
	LTS	Bias	1.4650	0.1921	0.0969	0.2202
		MSE	2.2362	0.0463	0.0173	0.0793
	LMS	Bias	0.9384	0.1761	0.2219	0.3415
		MSE	1.0656	0.0455	0.0657	0.3257
n =80	OLS	Bias	1.9789	0.1801	0.1203	0.1844
		MSE	3.9942	0.0428	0.0223	0.0474
	LAV	Bias	1.3392	0.1363	0.1095	0.2038
		MSE	1.8845	0.0277	0.0202	0.0717
	M-Huber	Bias	1.6679	0.1108	0.0727	0.1702
		MSE	2.8548	0.0245	0.0174	0.0448
	M-Turkey	Bias	1.4060	0.1706	0.0934	0.1567
		MSE	2.0780	0.0372	0.0183	0.0435
	LTS	Bias	1.4632	0.1718	0.1490	0.1665
		MSE	2.1798	0.0401	0.0340	0.0429
	LMS	Bias	0.8983	0.1404	0.2250	0.1698
		MSE	0.8930	0.0288	0.0967	0.0487
n =100	OLS	Bias	1.9481	0.1523	0.1187	0.1160
		MSE	3.8355	0.0294	0.0225	0.0207
	LAV	Bias	1.2564	0.0945	0.0857	0.1388
		MSE	1.6309	0.0162	0.0115	0.0342
	M-Huber	Bias	1.6520	0.0876	0.0825	0.1111
		MSE	2.7621	0.0117	0.0142	0.0183
	M-Turkey	Bias	1.3776	0.1100	0.0821	0.1111
		MSE	1.9417	0.0149	0.0130	0.0227
	LTS	Bias	1.4092	0.1065	0.1148	0.1262
		MSE	1.9995	0.0141	0.0183	0.0282
	LMS	Bias	0.7951	0.1509	0.1847	0.1000
		MSE	0.6746	0.0380	0.0673	0.0146

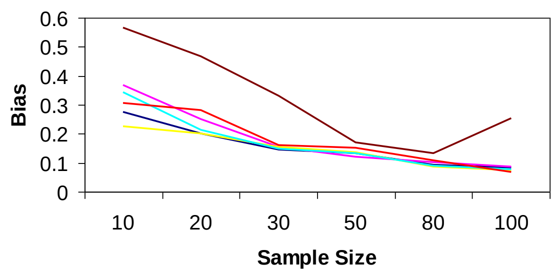
Source: Calculations based on simulated data

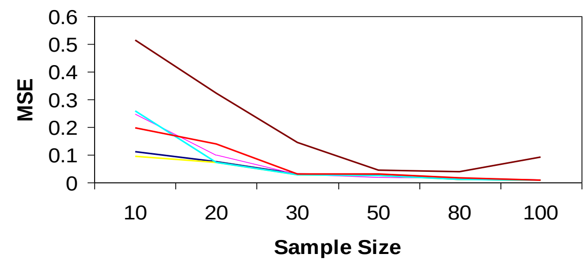




Intercept

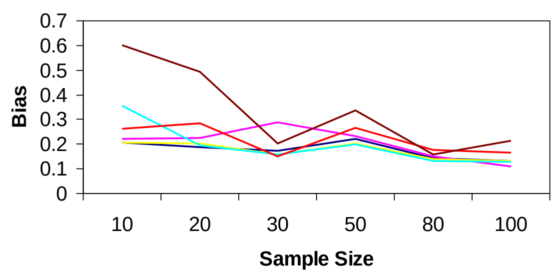
(a) (b)

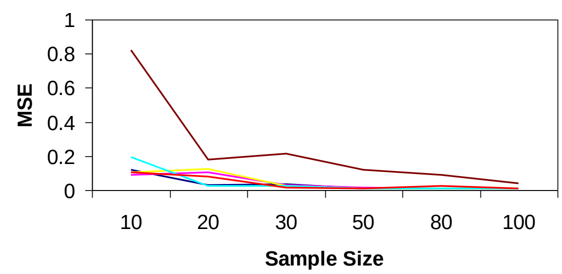




First Predictor

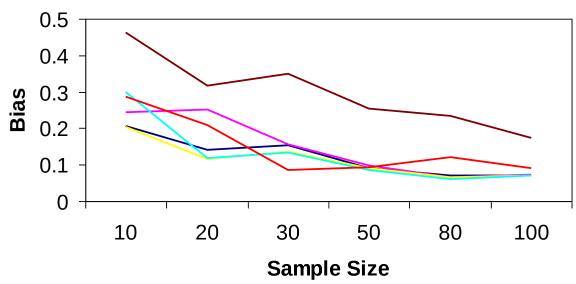
(c) (d)

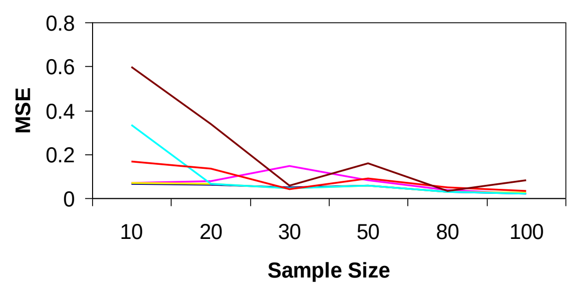




Second Predictor

(e) (f)





Third Predictor

(g) (h)

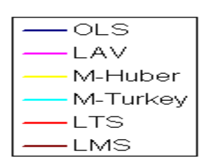
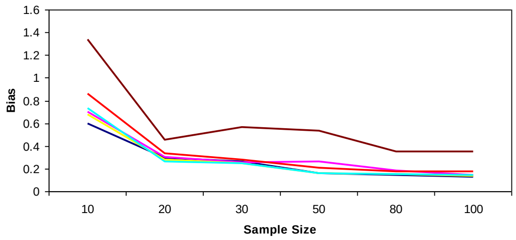
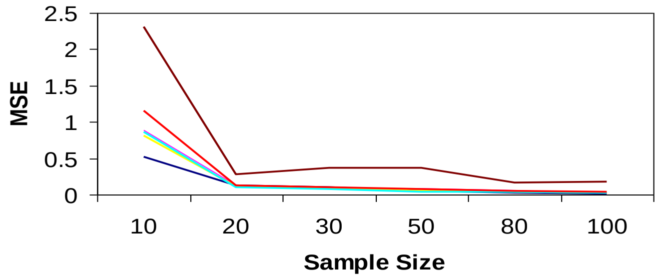


Figure 1 **Bias and MSE for 10 Simulations from Normal (0, 1) Distribution**

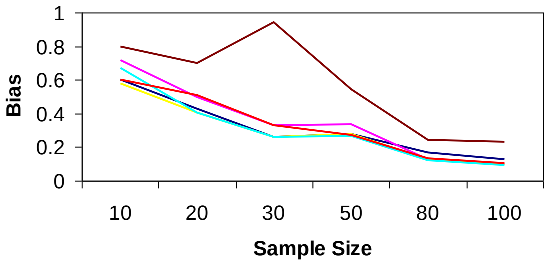
Source: **Table (2)**

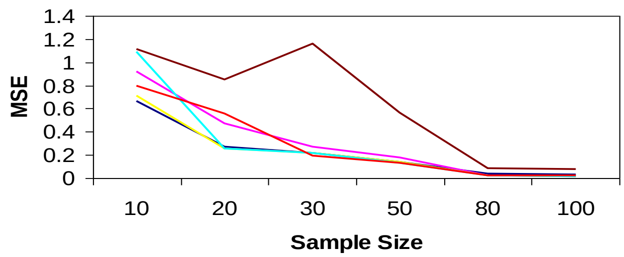




Intercept

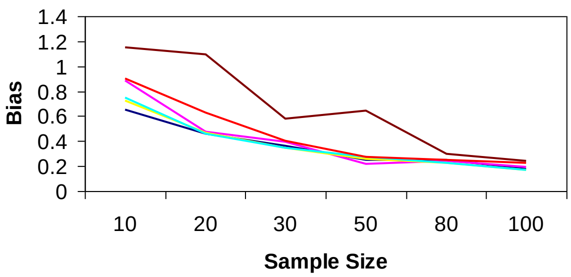
(a) (b)

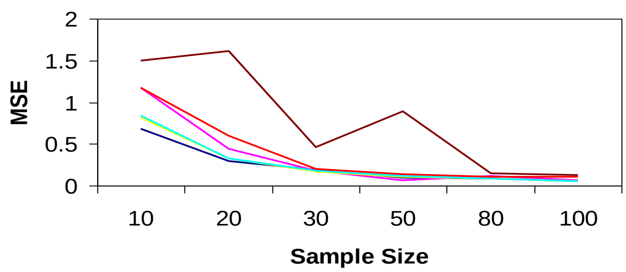




First Predictor

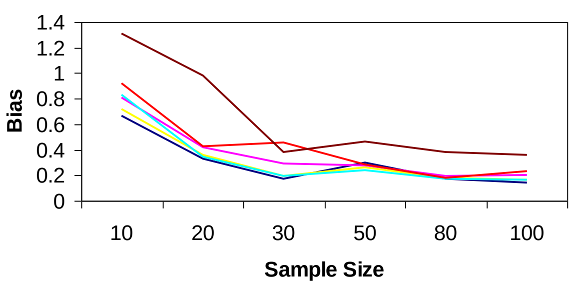
(c) (d)

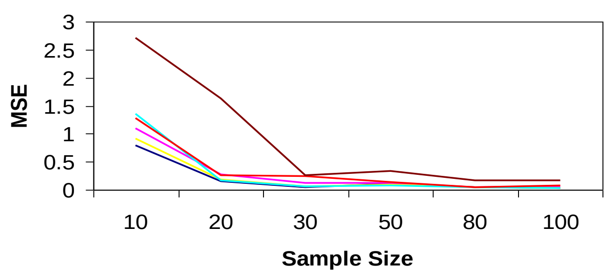


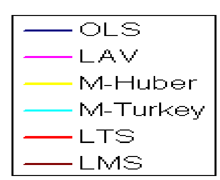


Second Predictor

(e) (f)



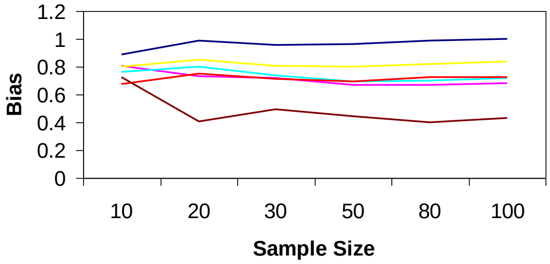


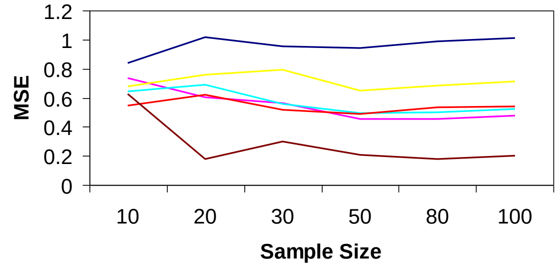


Third Predictor

(g) (h)

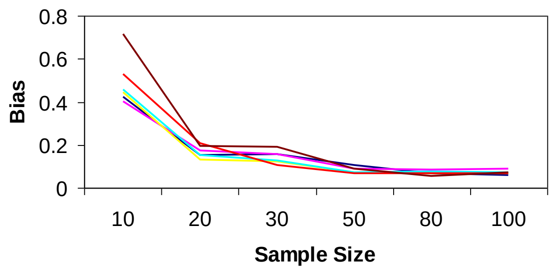
Figure 2 **Bias and MSE for 10 Simulations from Logistic (0, 1) Distribution**
Source: **Table (3)**

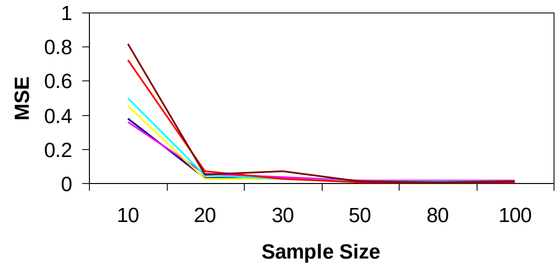




Intercept

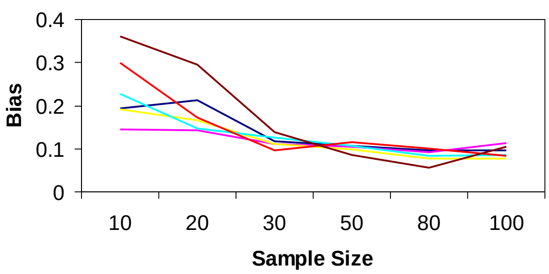
(a) (b)

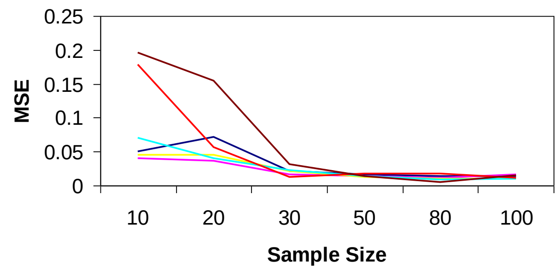




First Predictor

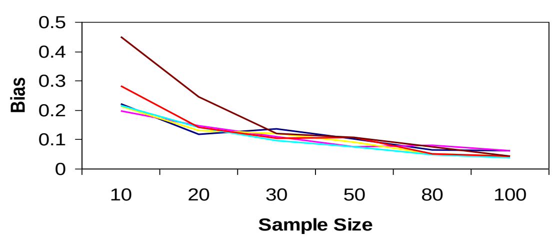
(c) (d)

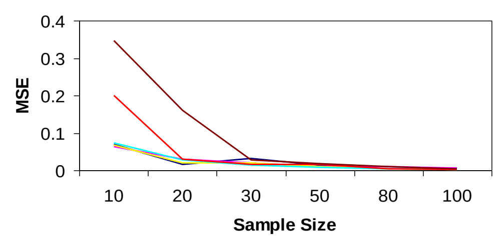




Second Predictor

(e) (f)





Third Predictor

(g) (h)

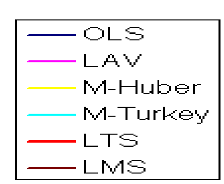
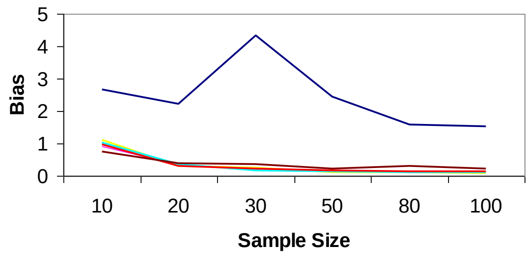
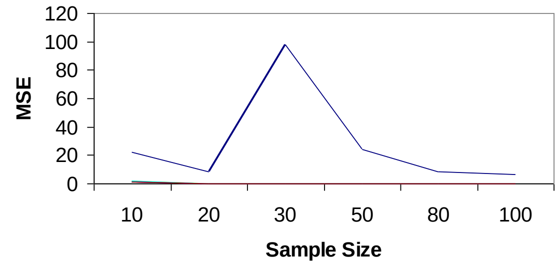


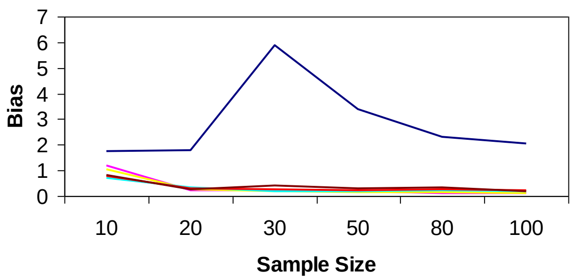
Figure 3 **Bias and MSE for 10 Simulations from Exponential (1) Distribution**
Source: **Table (4)**

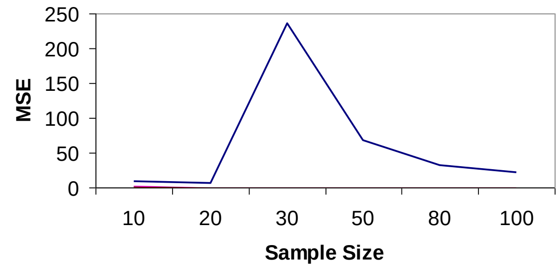




Intercept

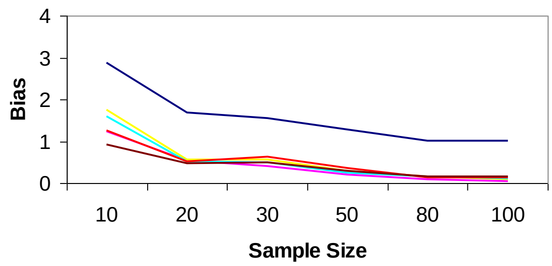
(a) (b)

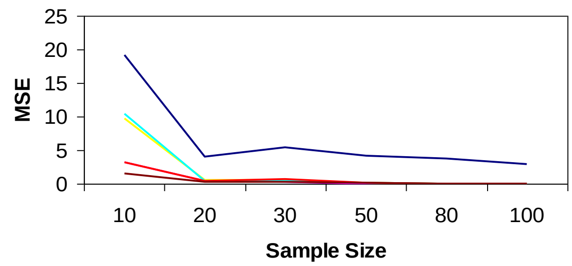




First Predictor

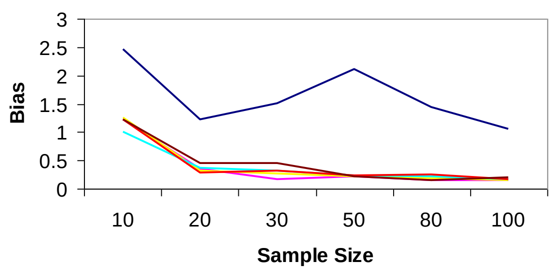
(c) (d)

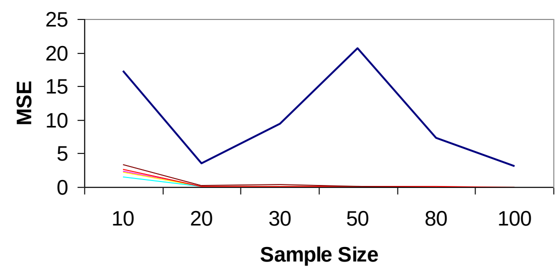




Second Predictor

(e) (f)





Third Predictor

(g) (h)

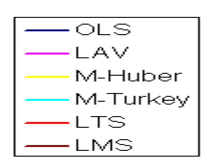
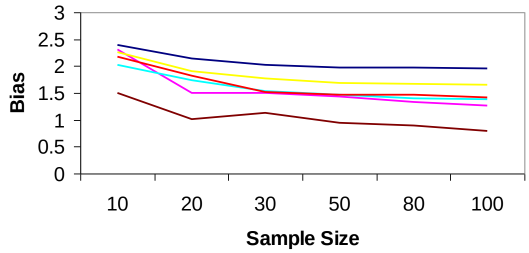
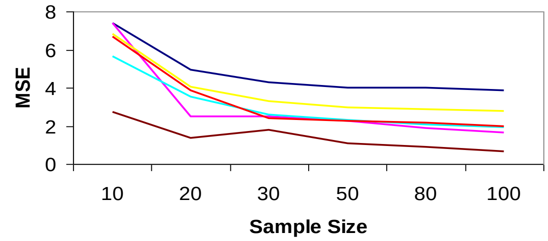


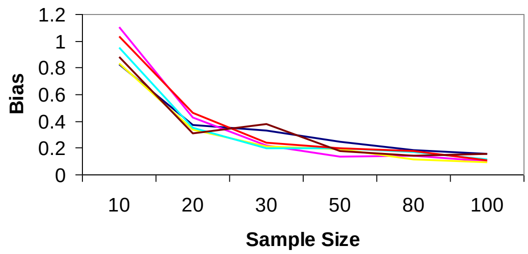
Figure 4 **Bias and MSE for 10 Simulations from Cauchy (0, 1) Distribution**
Source: **Table (5)**

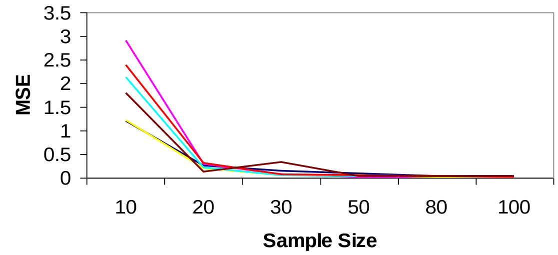




Intercept

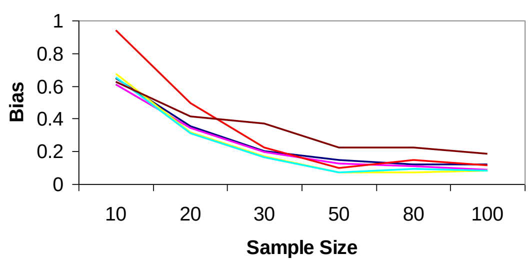
(a) (b)

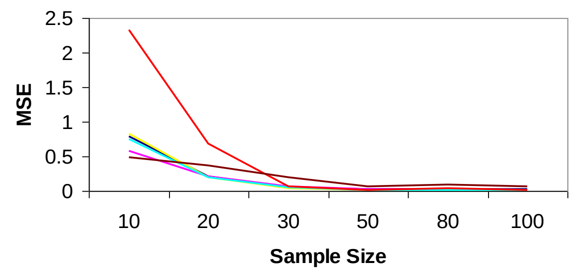




First Predictor

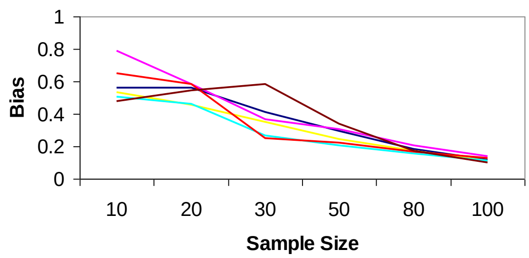
(c) (d)

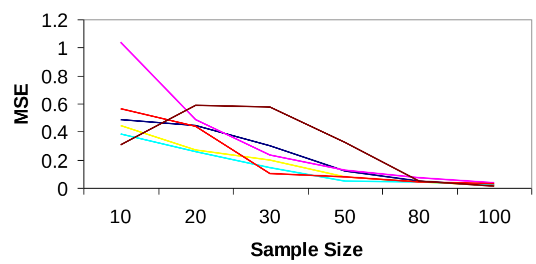




Second Predictor

(e) (f)





Third Predictor

(g) (h)

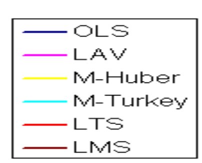


Figure 5 **Bias and MSE for 10 Simulations from Gamma (1, 0.5) Distribution**
Source: **Table (6)**

4. Conclusion

In order to study of distribution robustness in regression, the performances of six regression methods for two important classes of distributions namely symmetric and skewed are investigated. Different error structures such as normal, logistic, exponential, Cauchy and gamma distributions are used to find out the most suitable method. It is found that, the OLS method is more efficient than the robust methods under normal error distribution. In this case, the LMS method performs much worst.

In logistic distribution, the Turkey-*M* estimator is more robust than other estimation methods. Although no preferred robust method can be chosen in exponential and Cauchy distributions, the robust methods clearly outperform the OLS method. Moreover, it is shown that the OLS method performs much worst in the study of gamma distribution. The LMS method is more resistant in this distribution.

When outliers exist, the simulation results indicate that other alternatives of the OLS are more appropriate. Selecting a more efficient alternative to the OLS method is closely related to the type of data and so it is suitable to use several alternative methods in data analysis. In cases of skewed distributions, the performance of OLS is poorer as compared to other methods. Based on bias and MSE criteria, the LMS is more suitable for the exponential and gamma distributions.

In symmetric distributions investigated here, the MSEs are very close to one another for the sample sizes larger than 50 and so none of the estimation methods is superior in such circumstances. However, this is not true for the cases of the skewed distributions where the OLS method has shown to be far lower from the other methods of estimation. Compared to MSE criterion, the bias criterion fluctuated more and this fluctuation persists even for larger sample sizes. This instability of biases created some difficulties and confusion in finding the optimum estimation in some situations.

In summing up, when series are outlier contaminated (1%, 5% and 10%), an overall result is that outliers adversely affect the bias as well as MSE of OLS estimators. It is found that OLS estimation under a heavy-tailed distribution does not yield outlier robust estimates. Indeed, not only with the Gaussian distribution but also with the skewed distributions, OLS estimators failure in the presence of small levels of outlier contamination.

Acknowledgement

I'd like to thank to Dr. Tin Win, Rector of Yangon University of Economics, and Dr. Ni Lar Myint Htoo, Pro-rector of Yangon University of Economics, Who gave me a chance to participate in the Journal of Yangon University of Economics.

References

1. Draper, N. R., and H. Smith (1981), Applied *Regression Analysis*, New York: John Wiley.

2. Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W.A. Stahel (1986), *Robust Statistics*, *The Approach Based on Influence Functions,* New York: John Wiley and Sons.

3. Huber, P. J. (1964), Robust Estimation of a Location Parameter, Annals *of Mathematical Statistics*, vol. 35(1): 73-101.

4. Huber, P. J. (1973), Robust Regression: Asymptotic, Conjectures and Monte Carlo, *The Annals of Statistics,* vol. 1, 799-821.

5. Huber, P. J. (1981), *Robust Statistics,* New York: John Wiley and Sons.