International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 4299-4305

Modification of M-Test Using Geometric Median Covariance

Olusegun Kehinde Alo [1, 2] and Shamshuritawati Sharif [1]

[1]*School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara, 06010 UUM, Sintok Malaysia.*
Department of Mathematics and Statistics, Niger State Polytechnic, Zungeru, Nigeria.

[*]*Corresponding author*

Abstract

M-test which was first presented by Box (1949) is used to test the equality of several covariance matrices and very sensitive to departures from normality distribution. It is difficult to meet the assumption with the presence of outliers in data set. Since the estimation of sample mean vector and sample variance-covariance matrix will be inaccurate. Therefore, a modification of M-test that resolve the problem is required because not all kinds of data are competent to handle the inflexible of M-test assumptions. Hence, to overcome these drawbacks, we improved M-test by integrating the geometric median covariance for producing a new robust covariance equality test in the presence of outliers. It is known as M_{GMed}-test. Furthermore, we investigate the Type I error rate values for performance evaluation based on 5,000 simulation study. Several difference dimension: small, medium and large variables in relations to various sample sizes with conditions of rho, variance, epsilon, and kappa are used in estimating the performance of M-test and M_{GMed}-test in terms of Type I error rate values. Moreover, important results from this analysis are highlighted. Based on Type I error rate the performance of M_{GMed}-test is better than M-test.

Keywords: M-test, Covariance matrix, Type I error rate, Robust estimator and Outliers

INTRODUCTION

In multivariate analysis setting, collected data are contaminated by errors, with failure to fulfil the normality assumption can distort the Type I error rates [1]. This make distributional performance totally fails if the normality assumption is not completely met [2]. One problem that contribute to non-normality distribution is outliers. Outliers or contaminants are observations that are inconsistent in a data set [3, 4, 5,]. The presence of outliers usually can cause and harm the normality assumption [6].

Additionally, classical test rely on assumption that the data is normally distributed. If the assumption is violate, then the statistical result may not be appropriate. This test is highly sensitive to the presence of outliers which make the results inaccurate. It has damaging effects on statistical analysis, increasing the variance of error and reducing statistical test power [7, 8].

The method in which covariance matrix is estimated with the presence of outliers is key to multivariate analysis. Hence, the desired method to overcome the problem is by using robust scale and location estimator instead of the classical variance-covariance estimator.

Furthermore, variance-covariance matrix is compulsory in computing M-test. The M-test which was first presented by Box (1949) is used to test equality of several variance-covariance matrices. Conversely, M-test fails to incorporate outliers in the data sets.

As a result, robust estimators has been established as essential tools in analysing data that contain contaminated observation [9]. It can be used to identify outliers and provide resistant results when the presence of outliers.

Nevertheless, several types of robust estimators have been introduced and discussed in the past research.

Some of robust estimators multivariate location and dispersion are M-estimators [10], S-estimator [11], minimum covariance determinant (MCD), and minimum volume ellipsoid (MVE) [12], and MM-estimator [13]. More importantly, these estimators are resistance to outliers and it will produce the same result as the classical analysis, since it called for combination of high robustness and high efficiency values [14]. Therefore, a robust procedure that attempt to provide a good results will be used in this research work.

The early robust estimator for geometric median vector and sample covariance matrix was long presented in 1937, Weiszfeld introduced three proofs concerning the uniqueness of the geometric median in which one of the proofs provide an algorithm for its computations which is based on a strict convexity argument.

Recently, Efteliogu 2015, introduced the geometric median as a robust location estimator in statistics that can be applied to key domains such as, facility location

problem in strategic planning which aims to minimize the cost of transportation, in spatial economics where the land values are determined by the bid rent theory, and in biology and health domains where the outliers in high dimensional data are needed to be eliminated. The robustness properties of GMed are summarized under three important features, breakdown point, uniqueness and equivariance [15].

Moreover, GMed is an important concept in statistics where the location of the data sets is required regardless of the outliers present with 50% breakdown point, it is equivariant under Euclidean similarity transformations and it is unique under the condition that the points are not collinear, and the number of points in the set is odd [16].

In this study, the issue of outliers call for a modification of the existing statistical methods [17].

A new statistical test that able to solve the problems is required because not all kind of data competent to handle the inflexible of M-test assumptions. Therefore, we modify the M-test by integrating the geometric median covariance for developing a new robust covariance equality test in the presence of outliers.

METHODOLOGY

M-test is among the well-known test especially to the applied researchers. Mathematically, the original test [18] is given as follows,

$$M = \boldsymbol{\Upsilon} \sum_{i=1}^{m}(n_i - 1)\log\left|S_{ui}{}^{-1}\, S_u\right| \qquad (1.1)$$

where,

i. $\Upsilon = 1 - \frac{2p^2-3p-1}{6(p+1)(m-1)}\left(\sum_{i=1}^{m}\frac{1}{n_i-1} - \frac{1}{n-m}\right)$

ii. $S = \frac{(n_1 s_1 + n_2 s_2)}{n}$

iii. $S_u = \frac{n}{n-m}S$

iv. $S_{ui} = \frac{n_i}{n_i-1}S_i$

where, $\boldsymbol{S_u}$ and $\boldsymbol{S_{ui}}$ are the unbiased estimators of population covariance.

In order to develop robust M-test denoted by M_{GMed}-test the variance-covariance of geometric median, $S_{GMed(i)}$ where $i = 1, 2, \dots, m$ is substituted into equation 1.1.

Thus, the test now turn into the following equation,

$$M_{GMed} = \Upsilon \sum_{i=1}^{m}(n_i - 1)log\left|S_{M_{GMed(i)}}{}^{-1}\, S_{M_{GMed}}\right| \qquad (1.2)$$

where,

$$\Upsilon = 1 - \frac{2p^2+3p-1}{6(p+1)(m-1)}\left(\sum_{i=1}^{m}\frac{1}{n_i-1} - \frac{1}{n-m}\right)$$

$S_{M_{GMed(i)}}$ and $S_{M_{GMed}}$ are the i-th unbiased sample geometric median covariance estimator and the pooled population geometric median covariance matrix respectively with,

$$S_{M_{GMed}} = \frac{\sum_{i=1}^{m}(n_i - 1)S_{M_{GMed(i)}}}{n-m}$$

m is the number of subgroup where the stability of matrices is hypothesized.

$n_i = i$-th sample size. The measure of evaluation that is performed in M-test is the Type I error rate based on equation 1.1 and 1.2, are evaluated by using Type I error (α) rate with 5,000 simulation study.

RESULTS AND DISCUSSIONS

The major finding is on a new robust covariance equality test in the presence of outliers. We compared original M-test and modified M_{GMed}-test, in terms of Type I error rate values. For each of the test 4 types of data contamination are considered to examine the strength and weakness of the tests. Besides that, all these tests have been exposed to various conditions which are number of variables (p), sample sizes (n), percentages of outliers (epsilon) and probability used (Kappa). The comparisons are summarized in form of tables. The first column in each table shows the percentage of outliers (epsilon) and followed by shift of the mean (Kappa). The following six columns recorded the Type I error rate values of M_{GMed}-test and M-test, are investigated in the study. This conditions is repeated for different sample sizes.

The values that is closest to the significance level and within the 0.025 and 0.075 are shaded in the tables with yellow and blue colours, whereas, red colours denotes results cannot be produced. Additionally, Table 3.1 to 3.9 recorded the Type I error rate values for each condition are arranged based on the ascending number of variables, i.e. small ($p = 3, 5$ and 8), medium ($p = 10, 15$ and 20) and large number variable ($p = 25, 30$ and 50) with $\alpha = 0.05$.

Type I error for Small Number of Variables ($p = 3, 5$ and 8)

In Table 3.1 and 3.3 the Type I error rate of M_{GMed}-test and M-test, are recorded. The overall results show that M_{GMed}-test is more robust compared to M-test.

All the values of Type I error rate of M_{GMed}-test, fall within the interval [0.025, 0.075] in yellow colour when, $n = 10, 20, 30,$ and 50, at epsilon 0.1 and kappa, 5, except for the condition of $n = 100$, epsilon 0.05 and kappa, 2. On the other hand, we focus on M-test, in blue colour when, $n = 10, 20, 30, 50,$ and 100, perform well when epsilon 0.05 and kappa 2. From the result, it shows that M-test, still perform where only 61 out of 180 are robust and a total of 119 out of 180 are non-robust. Interestingly, the new statistical tests: M_{GMed}-test, performed very well more than twice the number, where 121 out of 180 are robust.

Additionally, Table 3.3 presented to show the Type I error of $p = 8$. The overall results shows that, M_{GMed}-test, is more robust with, 35 out of 60 are robust compared to M-test, where 37 out of 60 is non-robust. All the values of Type I error rates of M_{GMed}-test fall within the robustness range for, $n = 10$, except that 1 out of 10 conditions is non-robust. Moreover, the result of M_{GMed}-test, when, p = 8 is poor as compared to, p = 3 and 5.

In summary, there are 360 conditions involved in evaluating the robustness of test for small number of variables ($p = 3$, 5 and 8). There are 121 out of 360 conditions of M_{GMed}-test, and 61 conditions of M-test, fall within the robust interval.

Conversely, $p = 3$ and 5 there are 86 and 42 corresponding conditions of M_{GMed}-test, 19 and 19 corresponding to M-test, fall within the robust interval. In addition, when, $p = 8$, there are 58 out of 120 robust Type I error rate within the interval. As a conclusion, the Type I error rate for small number of variables show that M_{GMed}-test, is more robust when, $p = 3$ and 5 than, $p = 8$.

Table 3.1: Type I error rate for variable, p = 3

sample (n)		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2	0.047	0.054	0.053	0.060	0.042	0.055	0.040	0.050	0.039	0.055	0.034	0.060
	5	0.054	0.074	0.062	0.116	0.053	0.133	0.044	0.138	0.041	0.160	0.028	0.245
	10	0.064	0.113	0.074	0.229	0.065	0.310	0.056	0.364	0.053	0.409	0.040	0.629
0.1	2	0.049	0.058	0.051	0.068	0.037	0.068	0.033	0.061	0.031	0.067	0.023	0.086
	5	0.063	0.098	0.067	0.187	0.056	0.244	0.043	0.283	0.037	0.334	0.023	0.554
	10	0.085	0.187	0.092	0.402	0.082	0.559	0.070	0.646	0.067	0.724	0.054	0.929
0.2	2	0.050	0.061	0.049	0.087	0.030	0.094	0.024	0.092	0.019	0.109	0.010	0.185
	5	0.084	0.145	0.093	0.347	0.072	0.498	0.056	0.595	0.054	0.687	0.028	0.941
	10	0.132	0.330	0.152	0.690	0.133	0.863	0.120	0.922	0.118	0.967	0.096	0.999

Table 3.2: Type I error rate for variable, p = 5

sample (n)		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2	0.047	0.054	0.046	0.049	0.041	0.058	0.049	0.060	0.041	0.058	0.031	0.059
	5	0.053	0.070	0.059	0.101	0.050	0.142	0.060	0.339	0.051	0.184	0.031	0.291
	10	0.064	0.105	0.073	0.233	0.065	0.345	0.073	0.438	0.067	0.499	0.045	0.740
0.1	2	0.049	0.057	0.046	0.534	0.036	0.067	0.039	0.071	0.032	0.070	0.019	0.082
	5	0.064	0.098	0.073	0.180	0.061	0.267	0.064	0.339	0.052	0.390	0.029	0.651
	10	0.086	0.173	0.103	0.428	0.092	0.606	0.094	0.725	0.088	0.811	0.064	0.964
0.2	2	0.051	0.063	0.042	0.070	0.036	0.094	0.030	0.111	0.023	0.111	0.008	0.192
	5	0.076	0.138	0.107	0.345	0.100	0.547	0.090	0.673	0.079	0.780	0.047	0.973
	10	0.126	0.305	0.173	0.726	0.171	0.902	0.163	0.963	0.154	0.988	0.126	0.999

Table 3.3. Type I error rate for variable, p = 8

sample (n)		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2	0.047	0.049	0.062	0.059	0.045	0.054	0.046	0.062	0.044	0.059	0.037	0.070
	5	0.052	0.056	0.078	0.115	0.062	0.143	0.065	0.205	0.063	0.225	0.051	0.388
	10	0.057	0.069	0.097	0.248	0.079	0.370	0.083	0.510	0.081	0.572	0.072	0.832
0.1	2	0.048	0.050	0.060	0.062	0.045	0.063	0.042	0.075	0.041	0.075	0.025	0.105
	5	0.059	0.068	0.091	0.186	0.082	0.280	0.080	0.394	0.076	0.464	0.056	0.764
	10	0.070	0.098	0.131	0.444	0.118	0.662	0.118	0.800	0.116	0.869	0.100	0.991
0.2	2	0.052	0.052	0.056	0.444	0.045	0.087	0.035	0.109	0.032	0.075	0.014	0.234
	5	0.069	0.091	0.129	0.071	0.126	0.586	0.126	0.751	0.124	0.854	0.090	0.993
	10	0.095	0.160	0.219	0.749	0.209	0.930	0.214	0.980	0.208	0.995	0.178	1.000

Shaded region indicate Type I error within [0.025, 0.075]

Type I error for Medium Number of Variables $(p = \mathbf{10, 15}$ ***and*** $\mathbf{20})$

Table 3.4, 3.5 and 3.6 display Type I error rate for medium number of variables. Based on the results presented in three tables, the pattern of performance for $\mathrm{M_{GMed}}$-test, when $p = 10$, and 15 are equal there are both 52 out of 120 are robust, except when $p = 20$ there are 21 out of 120 are robust when $n = 10, 20, 30,$ 50 and 100 with 5% contamination and kappa 5 at least Type I error rate values.

From the simulation study, these three tables demonstrate that $\mathrm{M_{GMed}}$-test, is robust for at least all the sample sizes excluding 10% and 20% of data contamination. When $n = 10, 20, 30,$ and 50 with, $p = 10, 15$ and 20, M-test, is robust at 0% and 5% of data contamination with kappa 2. Meanwhile, when, $n = 100$, M-test, is robust at only 0% of data contamination under ideal condition (no contamination), but $n = 100$ failed when involved with large percentage of data contamination, 5%, 10% and 20%. Interestingly, M-test and $\mathrm{M_{GMed}}$ completely failed at all percentages of 0%, 5%, 10%, and 20% of data contamination, when, $n = 10$, $p = 10$; $n = 10$, $p = 15$; and $n = 10$ and 20, $p = 20$.

The simulation study for M-test, indicates that the values of Type I error fall within the robust interval under ideal condition (no contamination) only. When sample size increase, the number of condition that fall within the robust interval is decrease. At, $n = 100$, there are 2 conditions, and when sample size increase to 100, there are only 1 condition.

Meanwhile, for medium number of variables in all the sample sizes under consideration, all values of the Type I error rate for $\mathrm{M_{GMed}}$-test, for, $p = 10, 15$ and 20 are differ. For, $p = 10$, all the values are robust under 5% of data contamination at kappa 5.

For, $p = 15$, the number of condition that fall within the robust interval decreased whereas M-test, is robust under 5% contamination at kappa 2.

In summary, there are 360 conditions involved in evaluating the robustness of the test for medium number of variables ($p = 10, 15$ and 20). There are 73 out of 360 conditions of $\mathrm{M_{GMed}}$-test, and 38 conditions M-test, of that fall within the robust interval. Therefore, we concluded that $\mathrm{M_{GMed}}$-test, is more robust compared to all other tests.

Table 3.4. Type I error rate for variable, p = 10

sample $(\boldsymbol{n})$		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1			0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2			0.056	0.060	0.049	0.064	0.042	0.057	0.042	0.062	0.036	0.069
	5			0.069	0.107	0.066	0.160	0.065	0.073	0.061	0.249	0.049	0.416
	10			0.089	0.210	0.089	0.401	0.085	0.492	0.082	0.598	0.068	0.857
0.1	2			0.054	0.062	0.051	0.069	0.043	0.198	0.041	0.079	0.025	0.100
	5			0.082	0.175	0.090	0.305	0.083	0.407	0.818	0.498	0.061	0.791
	10			0.124	0.397	0.133	0.677	0.126	0.806	0.125	0.889	0.104	0.994
0.2	2			0.056	0.075	0.050	0.097	0.041	0.110	0.036	0.129	0.016	0.234
	5			0.120	0.338	0.137	0.611	0.141	0.771	0.129	0.869	0.113	0.993
	10			0.215	0.707	0.228	0.940	0.227	0.986	0.219	0.995	0.204	1.000

Table 3.5: Type I error rate for variable, p = 15

sample $(\boldsymbol{n})$		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1			0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2			0.047	0.048	0.050	0.061	0.050	0.061	0.048	0.068	0.040	0.076
	5			0.053	0.078	0.068	0.145	0.068	0.205	0.069	0.267	0.059	0.495
	10			0.073	0.137	0.093	0.345	0.092	0.495	0.090	0.626	0.078	0.907
0.1	2			0.049	0.053	0.053	0.073	0.051	0.075	0.045	0.082	0.033	0.115
	5			0.066	0.120	0.090	0.290	0.093	0.413	0.091	0.544	0.078	0.858
	10			0.105	0.264	0.140	0.630	0.140	0.809	0.138	0.908	0.118	0.996
0.2	2			0.051	0.060	0.058	0.097	0.048	0.108	0.053	0.137	0.028	0.254
	5			0.090	0.225	0.147	0.582	0.157	0.787	0.169	0.901	0.159	0.998
	10			0.173	0.264	0.256	0.927	0.255	0.985	0.263	0.997	0.255	1.000

Table 3.6: Type I error rate for variable, p = 20

sample (n)		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1					0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2					0.035	0.038	0.052	0.065	0.050	0.067	0.044	0.070
	5					0.046	0.095	0.072	0.203	0.071	0.277	0.065	0.515
	10					0.071	0.232	0.102	0.480	0.095	0.616	0.084	0.913
0.1	2					0.038	0.044	0.054	0.079	0.053	0.083	0.041	0.112
	5					0.060	0.195	0.096	0.417	0.099	0.557	0.092	0.879
	10					0.108	0.499	0.146	0.799	0.144	0.903	0.131	0.998
0.2	2					0.043	0.068	0.065	0.122	0.057	0.138	0.049	0.260
	5					0.108	0.456	0.174	0.778	0.173	0.903	0.199	0.998
	10					0.209	0.862	0.278	0.983	0.272	0.999	0.281	1.000

Shaded region indicate Type I error within [0.025, 0.075]

Type I error rate for Large Number of Variables **(*p* = 25, 30 *and* 50)**

Type I error rate of M-test, and M_{GMed}-test, for large number of variables are presented in Table 3.7, 3.8 and 3.9. Based on the three tables, for all combination of sample sizes, M_{GMed}-test, is robust under ideal condition (no contamination) and 5% of data contamination.

Meanwhile, for M-test, the Type I error fall within the robust interval for under ideal condition (no contamination) only as the sample sizes increase from, $n = 40, 50$ and 100. Overall, when, $p = 25, 30$ and 50, M_{GMed}-test, has 21, 13 and 3 out of 360 conditions that fall within robust interval, correspondingly. Therefore, we summarized that M_{GMed}-test, is more powerfully robust in large number of variables.

Table 3.7: Type I error rate for variable, p = 25

sample (n)		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1					0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
0.05	2					0.054	0.061	0.051	0.063	0.048	0.068	0.043	0.075
	5					0.063	0.108	0.063	0.183	0.067	0.251	0.069	0.550
	10					0.082	0.206	0.090	0.412	0.092	0.577	0.088	0.919
0.1	2					0.056	0.067	0.054	0.079	0.049	0.087	0.042	0.116
	5					0.075	0.188	0.088	0.374	0.087	0.529	0.144	0.900
	10					0.114	0.400	0.141	0.738	0.139	0.892	0.305	1.000
0.2	2					0.059	0.076	0.059	0.108	0.058	0.133	0.054	0.265
	5					0.108	0.369	0.143	0.752	0.165	0.898	0.305	1.000
	10					0.205	0.761	0.253	0.976	0.272	0.999	0.305	1.000

Table 3.8: Type I error rate for variable, p = 30

***sample* (n)**		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1							0.050	0.050	0.050	0.050	0.05	0.050
0.05	2							0.049	0.061	0.060	0.074	0.052	0.085
	5							0.061	0.160	0.077	0.240	0.081	0.579
	10							0.087	0.343	0.104	0.534	0.100	0.924
0.1	2							0.054	0.075	0.059	0.087	0.053	0.133
	5							0.080	0.321	0.092	0.497	0.124	0.919
	10							0.133	0.660	0.149	0.865	0.165	0.998
0.2	2							0.063	0.105	0.071	0.135	0.067	0.288
	5							0.131	0.649	0.168	0.870	0.250	1.000
	10							0.248	0.951	0.279	0.997	0.333	1.000

Table 3.9: Type I error rate for variable, p = 50

***sample* (n)**		n=10		n=20		n=30		n=40		n=50		n=100	
epsilon	***kappa***	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M	M_{GMed}	M
0	1									0.050			0.050
0.05	2									0.044			0.942
	5									0.088			0.590
	10									0.095			0.907
0.1	2									0.041			0.152
	5									0.153			0.929
	10									0.179			0.998
0.2	2									0.083			0.314
	5									0.440			1.000
	10									0.534			1.000

Shaded region indicate Type I error within [0.025, 0.075]

CONCLUSION

The ultimate goal of this study is to develop a new robust method of M-test in testing and applying the new statistical covariance equality test in the presence of outliers. In achieving the objective, the integration of the M-test into robust location estimators of M-geometric median test by using M-estimator and M_{GMed}-estimator. The test is use to handle the problem of outliers. Moreover, the Type I error rate value is conducted to evaluate the performance of M-test and M_{GMed}-test using simulated data. Generally, the results showed that, M_{GMed}-test dominated M-test for every different number of variables except when $p = 5$. The robustness of M-test differs among small, medium and large number of variables. In small variable, when the sample sizes increase, the performance of M-test becomes worst in condition of high percentage (10%, 15% and 20%) of data contamination. While, medium variable, $n = 20, 30$ and 40, M-test is robust under idea condition (no contamination) and 5% of data contamination. From the results it showed that M-test performed well in medium variable and worst at 20%. Meanwhile, M-test, the Type I error fall within the robust interval for under ideal condition (no contamination) only.

ACKNOWLEDGEMENT

We the authors appreciate the financial support from the Universiti Utara Malaysia for carrying out this research. Also we thank the reviewers that take their time to perfect this article.

REFERENCE

[1] Yusof, Z. M., Haruna, N. H., Syed Yahaya, S. S. Abdullah, S., *Modified parametric bootstrap: A robust alternative to classical test.* In: World Conference on Integration of Knowledge: Business, Economy, Management, Cultural, Societal, Theological and Scientific (WCIK 2013), 25 & 26 November 2013, Langkawi, Malaysia.

[2] Aslam, S., & Rocke, D. M. (2005). A robust testing procedure for the equality of covariance matrices. *Journal for Computational Statistics & Data Analysis*, *49*(3), 863-874.

[3] Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (Vol. 3, No. 1). New York: Wiley.

[4] Yuan, K. H., and Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. The British Journal of Mathematical and Statistical Psychology, 54,161-175. doi:10.1348/000711001159366.

[5] Sharif, S., Djauari, M.A., & Djauari, H. (2012). Correlation network analysis on safety culture and worker's behaviour and safety culture: an experience in manufacturing industry. International Journal of Basic & Applied Sciences, 12(5), 1-6.

[6] Ali, H. B., (2013), "Efficient and highly robust Hotelling T^2 Control chart using Reweighted minimum vector variance", "A Ph.D thesis submitted to Universiti Utara Malaysia".

[7] Osborne, J. W., & Overbay, A. (2012). Best practices in data cleaning. Copyright ©2013 by *SAGE Publications, Inc., Printed in the United States of America* ISBN 978-1-4129-8801-8(PbK).

[8] Dan, E., D. and Ogu, A., I. (2013). Statistical analysis/methods of detecting outliers in a univariate data in a regression analysis model. *International Journal of Education and Research Vol. 1 No. 5 ISSN: 2201-6333 (Print) ISSN: 2201-6740 (Online) www.ijern.com.*

[9] Muthukrishnan, R., and Radha, M., (2010). "M-estimator in regression models", Journal of Mathematics Research, 2(4), 23-27.

[10] Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73- 101.

[11] Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 1269-1292.

[12] Rousseeuw, P. J., (1985). Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, *8*, 283-297.

[13] Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 642-656.

[14] Roelant, E., Van Aelst, S., & Croux, C. (2009). Multivariate generalized S-estimators. *Journal of Multivariate Analysis*, *100*(5), 876- 887.

[15] Eftelioglu, E., (2015). Geometric median. Computer science and Engineering, university of Minnesota, Minneapolis, MN, USA. Springer International Publishing Switzerland. S. Shekhar et al. (eds.), Encylopedia of GIS, DOI 10.1007/978-3- 319- 23519-6_1555-1.

[16] Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. Biometrika 35, 414.

[17] Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review*, *1*(2), 293-314.

[18] Box, G., E., P., (1949). A general distribution theory for a class of likelihood criteria. *Biometrika,* 36: 317-346.