Applied Mathematical Sciences, Vol. 11, 2017, no. 20, 953 - 961
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2017.7276

Characteristics of Group LASSO in Handling High Correlated Data

M. Yunus

Department of Statistics, Faculty of Mathematics and Natural Science
Bogor Agricultural University, Indonesia

Asep Saefuddin and Agus M. Soleh*

Department of Statistics, Faculty of Mathematics and Natural Science
Bogor Agricultural University, Indonesia
*Corresponding author

Abstract

Problems of high correlated data in a linear regression can not be handled directly by standard methods of parameter estimation such as the least squares (LS). Lasso technique is a proper method to handle this problems. However, lasso may remove some important covariates. Group lasso technique is an extension of the lasso which allegedly be able to overcome it. The study was conducted by creating a simulation to study the characteristics of the group lasso. The simulation results showed the group lasso was better than LS and lasso technique when group of β is present included in a group. In other case, group lasso remained the same as LS and lasso.

Keywords: Correlated data, Group lasso, Lasso, Least squares, Simulation

1. Introduction

Problems of high correlated data in a linear regression can not be handled directly by the least squares (LS) method. One solution can be used to overcome this problem is subset selection techniques (such as best subsets, selection forward, backward elimination, stepwise), shrinkage (such as Ridge Regression), and the transformation of dimension (such as PCA and PLS). In general, subset selection techniques provide an advantage in the selection of important covariates but its prediction is unstable. On the other hand, the technique of shrinkage and the transformation of dimension give the opposite of the subset selection techniques. Technique of least absolute shrinkage and selection operator (lasso) was first introduced [8] that have both advantages (stable and important covariate information). Lasso technique transforms a penalty in ridge regression with the norm L_1 (regularization of L_1) and started known after [1] found LAR (Least Angle Regression) algorithm method. The lasso technique gave better results for correlated data modelling than LS [7]. The weakness of this technique is still possible to eliminate some important covariates that have a high correlation and a tendency to be over estimated [6]. Group lasso technique is an extension of the lasso which allegedly be able to overcome it. Group lasso technique was introduced by Bakin in 1999 and Lin and Zhang in 2006, and further generalized by Yuan and Lin in 2007 [2].

Some studies on group lasso had been conducted (see: [10] (model selection and prediction of the variables that have been grouped), [5] (estimation of group lasso for a linear model), and [9] (identification of consistency of group lasso estimation results)). However, the characteristics of the group lasso to overcome a high correlated data has still unknown. Therefore, this study will examine the characteristics of the group lasso on the high correlated data.

2. Group LASSO

If there is an input vector of $\boldsymbol{X}^T = (X_1, X_2, \dots, X_p)$ and used to predict the outcomes in the form of real number, then the linear regression model has the following form:

$$y_i = \beta_0 + \sum_{j=1}^{p} X_{ij}\,\beta_j + \varepsilon_i$$

The parameter β is estimated by using the least squares method, which minimizes the sum of squared residual [2]. At the high correlated data, one of the estimation methods is lasso technique that its function can be written in the form of Lagrange's equation, as follows:

$$\underset{\beta}{\operatorname{argmin}} \left\{ \left\| \mathbf{y} - \beta_0 \mathbf{1} - \sum_{i=1}^{p} \mathbf{X}_l \beta_i \right\|_2^2 + \lambda \, \|\boldsymbol{\beta}\|_1^1 \right\}, \qquad \lambda \geq 0.$$

in this case, λ is the lasso parameter with the value of $\lambda \geq 0$.

Group lasso allows grouped variables [3] and gives better results in predicting compared to lasso [4]. Estimator of coefficients in group lasso is obtained by minimizing the sum of squared residuals with the constraint is similar to the lasso but some covariates are grouped [10]. Suppose that the p covariates are divided into l groups, with p_l the number in group l, $\boldsymbol{X}_l$ represent covariates corresponding to the l^{th} group, with corresponding coefficient vector $\boldsymbol{\beta}_l$.

$$\underset{\beta}{\operatorname{argmin}}\left(\left\|\mathbf{y}-\beta_0\mathbf{1}-\sum_{l=1}^{L}\mathbf{X}_l\beta_l\right\|_2^2+\lambda\sum_{l=1}^{L}\sqrt{p_l}\,\|\boldsymbol{\beta}_l\|_2^2\right)$$

Where the $\sqrt{p_l}$ terms accounts varying group sizes and $\|.\|_2$ is the Euclidean norm (not squared). While $\lambda \geq 0$ as tuning parameters of shrinkage.

3. Method

Data

The data used in this study was the data from simulation which consisted of six covariates and one response. Covariates generated from normal multivariate with correlation of 0.9 between all of X_i and X_j pairs. The response generated from a linear model with seven scenarios. The generation procedures in more detail are presented in the procedure section.

Procedures

Procedures of simulation data analysis for studying the characteristic of group lasso are as follows:

a. Generate covariates $(X_1,X_2,\dots,X_6,)$ with each mean is 1 $(\mu_i = 1)$ and the correlation is 0.9 $(\rho = 0.9)$.
b. Generate response variable (Y) with several coefficient scenarios $(\beta_1,\beta_2,..,\beta_6)$ are the combination of $\beta_i = 0$, $0 < \beta_i < 1$, and $\beta_i \geq 1$. All the standard error were normal distribution $(\varepsilon \sim N(0,\sigma))$. The scenarios are presented in Table 1.

Table 1. The Scenario β_i on Data Simulation

No	β_0	Group 1			Group 2		Group 3
		β_1	β_2	β_3	β_4	β_5	β_6
1.	1	0.7	0.7	0.7	0.7	0.7	0.7
2.	1	2.0	2.0	2.0	2.0	2.0	2.0
3.	1	0.7	0.7	0.7	0	0	0.7
4.	1	0.7	0	0.7	0	0	0.7
5.	1	0.7	0.7	2.0	0	0	0.7
6.	1	−0.7	0.7	0.7	0.7	0.7	0.7
7.	1	0.7	0.7	0.7	−0.7	−0.7	0.7

c. Model estimation is performed by:
    i. Least square (LS) method
    ii. Least absolute shrinkage and selection operator method
    iii. Group least absolute shrinkage and selection operator method
d. Procedure a – c repeated 100 times
e. The results were presented in *boxplot* form on one graphic

4. Result and Discussion

The simulation results are presented in Figure 1, Appendix 1 - Appendix 6. The result of simulation estimation generally using the least squares (LS) method resulted in a greater variance compared to lasso and group lasso technique. When the actual parameter coefficients equal to zero and included in the same group, lasso and group lasso can select variables properly. These methods provide more homogeneous distribution on a regression coefficient parameters which have a value of zero $(\beta_i = 0, i = 1, \ldots, 6)$. The result of parameter coefficient estimation was obtained relatively similar between lasso and groups lasso and, even though there was tend that estimation variance using group lasso was smaller than lasso.

The results of the first simulation scenario $(\beta_i = (0.7, 0.7, 0.7, 0.7, 0.7, 0.7)^t)$ is presented in Figure 1, while the second $(\beta_i = (2, 2, 2, 2, 2, 2)^t)$ and the third $(\beta_i = (0.7, 0.7, 0.7, 0, 0, 0.7)^t)$ scenario are presented in Appendix 1 and Appendix 2, respectively. The technique of group lasso provided the smaller variance of coefficient than the other two methods, however lasso and group lasso techniques had a tendency to underestimate (seen from coefficient parameter estimation, more than 50% of the estimation result were smaller than the actual value). In this simulation, group lasso could estimate a better regression coefficient, which gave the homogeneous distribution on the parameters of regression coefficients that were grouped before. However, the estimation that was obtained by lasso and group lasso technique tended to be smaller than coefficient estimation using LS method.

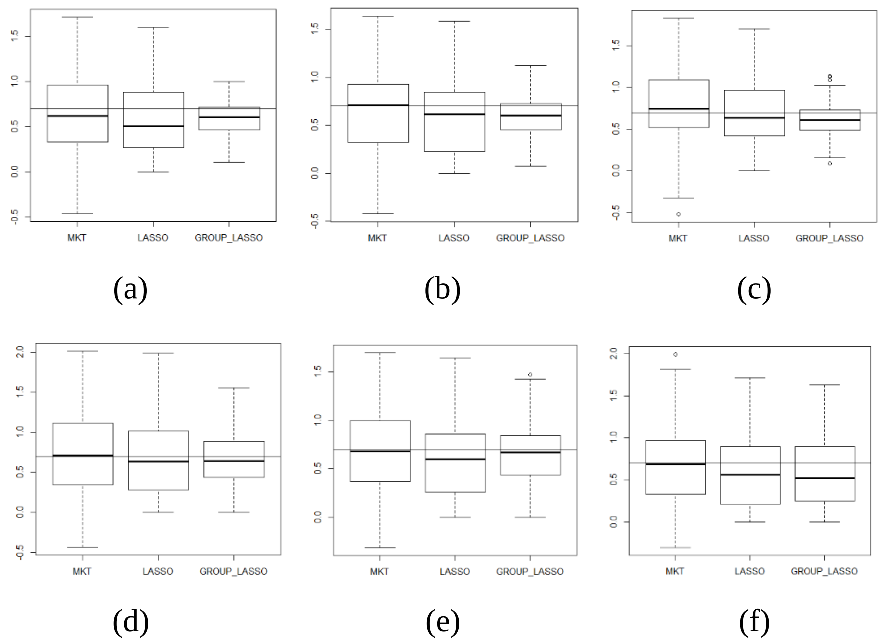


Figure 1. The plot of 100 times simulation estimation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (0.7, 0.7, 0.7, 0.7, 0.7, 0.7)^t$

The result of the fourth simulation scenario $(\beta_i = (0.7, 0, 0.7, 0, 0, 0.7)^t)$ and the fifth $(\beta_i = (0.7, 0.7, 2, 0, 0, 0.7)^t)$ are presented in Appendix 3 and Appendix 4. Estimation of parameter coefficient and selection of variables that was performed by lasso technique was better than LS and group lasso technique, in which the variance of estimated coefficient which generated by using LS was higher than the other two methods. Group lasso technique had a tendency to underestimate at $\beta_i > 0$ (seen in estimation of coefficient parameter which was almost 100% of estimation was smaller than the actual value) and over estimated at $\beta_i = 0$ (seen in estimation of coefficient parameter which was supposed to be selected). In lasso technique, it also had a tendency to underestimate. In this simulation, when grouping was done, there was still insignificant variables in its group. However, group lasso is not better than LS and lasso.

The scenario results of the sixth $(\beta_i = (-0.7, 0.7, 0.7, 0.7, 0.7, 0.7)^t)$, and seventh $(\beta_i = (0.7,0.7,0.7,-0.7,-0.7,0.7)^t)$ simulation are presented in Appendix 5 and Appendix 6, respectively. Estimation using LS generated a greater variance compared to lasso and group lasso techniques. In group lasso technique, it had a tendency to under and over estimate. In lasso technique, it also had a tendency to under estimate. Several tendencies found in these simulations are as follow; firstly when grouping was done, and there were variables that possess a high effect $(\beta_i \geq 1)$ while the other had a low effect $(0 < \beta_i < 1)$, the second when the variables had a positive or negative effect in a group, and the third when the effect between groups is different namely positive and negative. Group lasso was not better than LS and lasso technique. However, for the group selection, group lasso is better than the other two methods.

5. Conclusion

Lasso and group lasso provide an alternative for selecting variables and estimation of regression coefficients on a high correlated covariates. Group lasso is better than LS and lasso in the presence of group β and when included in one group, $\beta_i \geq 1$ all, $\beta_i < 1$ all, or $\beta_i = 0$ all. In another case, group lasso is not better than lasso and LS. However, it can be used as an alternative solution.

References

[1] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Stat.*, **32** (2004), 407-499. https://doi.org/10.1214/009053604000000067

[2] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York, 2009. https://doi.org/10.1007/978-0-387-84858-7

[3] J. Huang, T. Zhang, The benefit of group sparsity, *Ann. Stat.*, **38** (2010), 1978-2004. https://doi.org/10.1214/09-aos778

[4] K. Lounici, M. Pontil, S. Van De Geer and A. B. Tsybakov, Oracle inequalities and optimal inference under group sparsity, *Ann. Stat.*, **39** (2011), 2164-2204. https://doi.org/10.1214/11-aos896

[5] Y. Nardi, A. Rinaldo, On the asymptotic properties of the group lasso estimator for linear models, *Electron. J. Stat.*, **2** (2008), 605-633. https://doi.org/10.1214/08-ejs200

[6] A.M. Soleh, Aunuddin, LASSO: Solusi alternatif seleksi peubah dan penyusutan koefisien model regresi linier, *Forum Statistika dan Komputasi*, **18** (2013), 21-27.

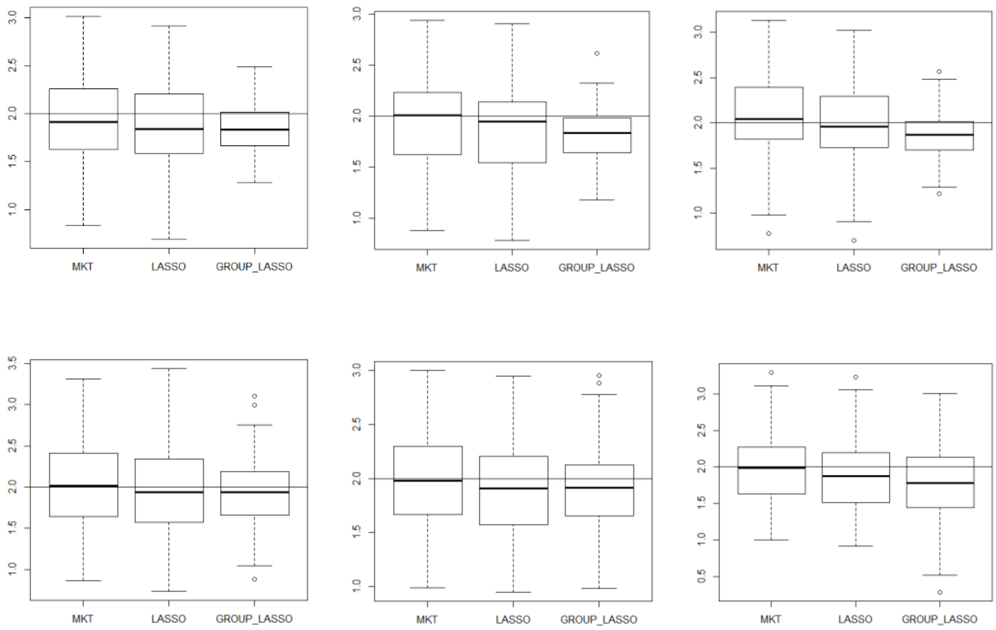
[7] A. M. Soleh, A. H. Wigena, A. Djuraidah, A. Saefuddin, Statistical downscaling to predict monthly rainfall using linear regression with L_1 regularization (lasso), *Applied Mathematical Sciences*, **9** (2015), 5361-5369. https://doi.org/10.12988/ams.2015.56434

[8] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. Series B*, **58** (1996), 267-288.

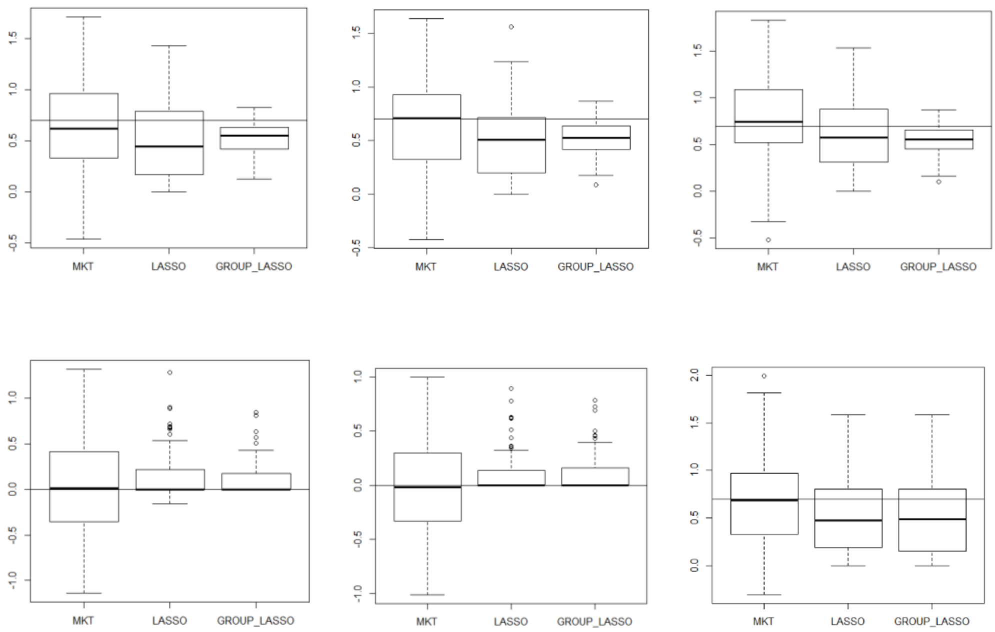
[9] H. Wang, C. Leng, A note on adaptive group lasso, *Comput. Stat. Data Anal.*, **52** (2008), 5277-5286. https://doi.org/10.1016/j.csda.2008.05.006

[10] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc. Series B*, **68** (2006), 49-67. https://doi.org/10.1111/j.1467-9868.2005.00532.x

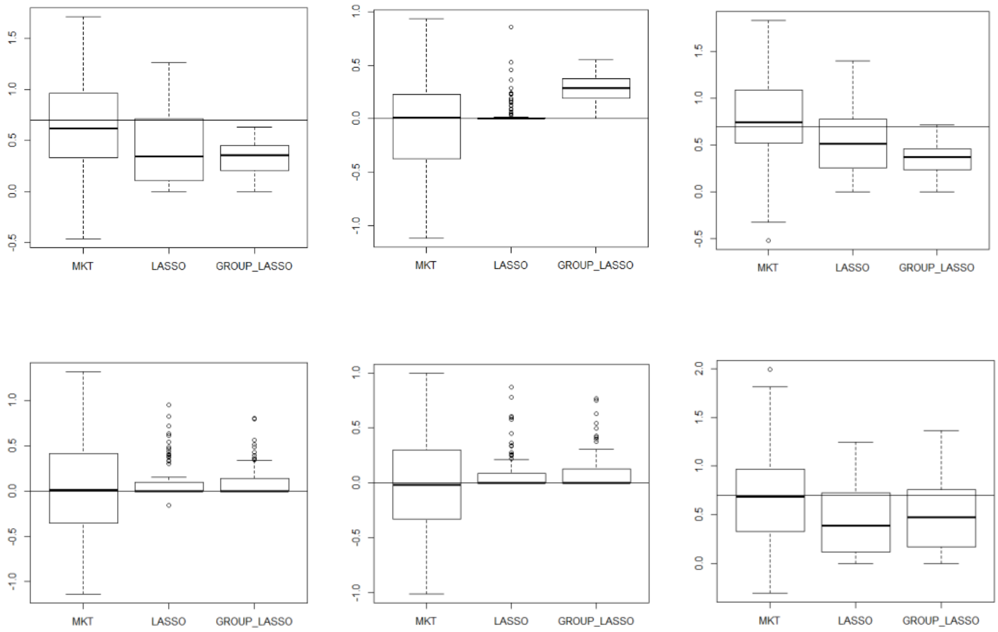
Appendix 1. The plot of 100 times of estimation simulation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (2, 2, 2, 2, 2, 2)^t$



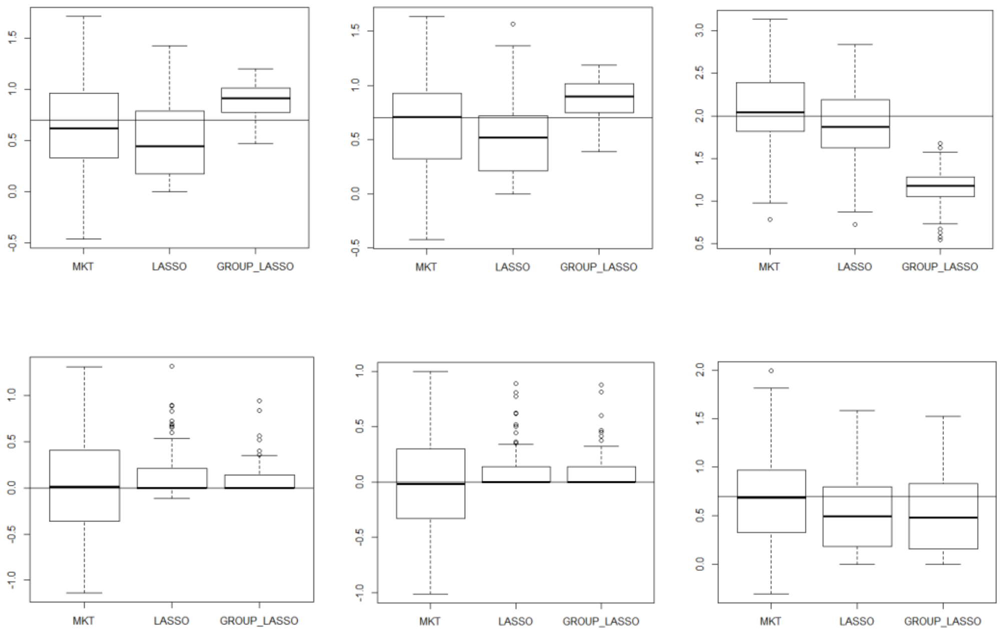
Appendix 2. The plot of 100 times of estimation simulation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (0.7, 0.7, 0.7, 0, 0, 0.7)^t$



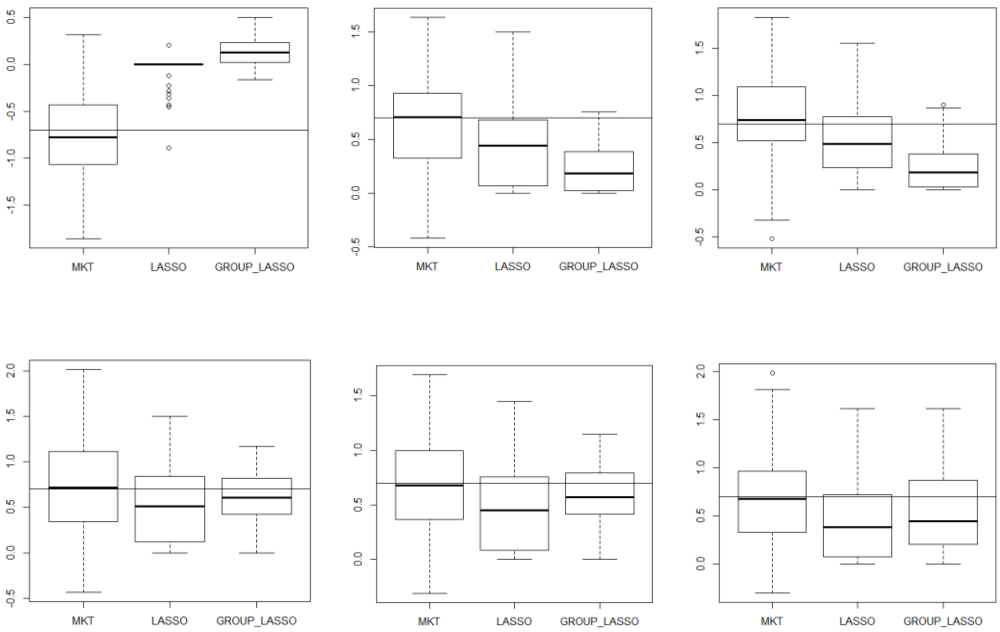
Appendix 3. The plot of 100 times of estimation simulation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (0.7, 0, 0.7, 0, 0, 0.7)^t$



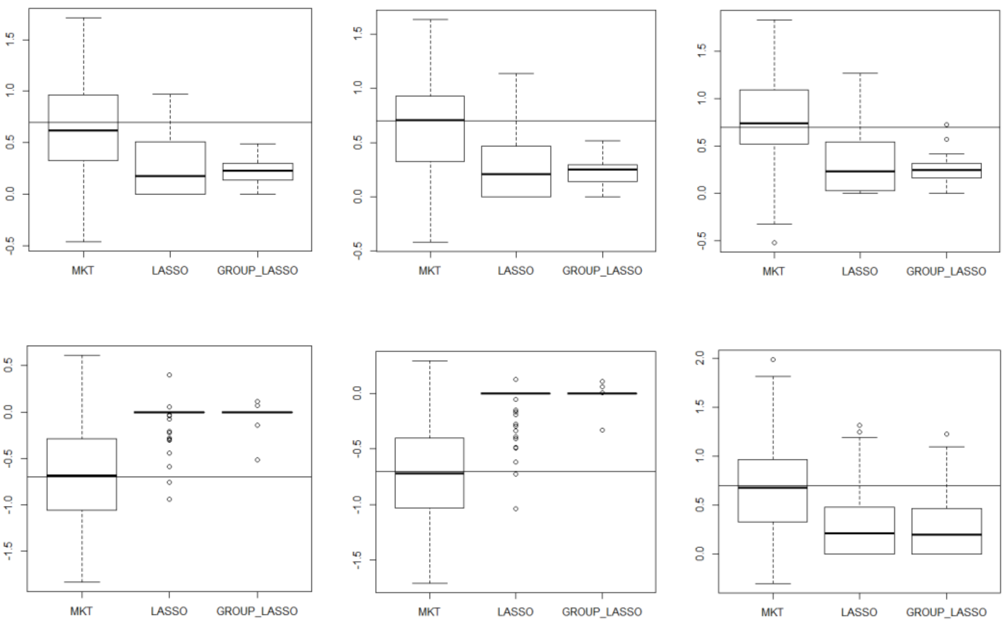
Appendix 4. The plot of 100 times of estimation simulation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (0.7, 0.7, 2, 0, 0, 0.7)^t$



Appendix 5. The plot of 100 times of estimation simulation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (-0.7, 0.7, 0.7, 0.7, 0.7, 0.7)^t$



Appendix 6. The plot of 100 times of estimation simulation with a correlation of $\rho = 0.9$ for coefficient $\beta_i = (0.7, 0.7, 0.7, -0.7, -0.7, 0.7)^t$



Received: March 9, 2017; Published: March 31, 2017