Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 6, Issue, 01, pp.4505-4510, January, 2014

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

# COMBINED ESTIMATORS AS ALTERNATIVE TO MULTICOLLINEARITY ESTIMATION METHODS

*Kayode Ayinde and Adewale F. Lukman

Department of Statistics, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Oyo State, Nigeria

ARTICLE INFO

*Article History:*
Received 24th September, 2013
Received in revised form
11th October, 2013
Accepted 19th December, 2013
Published online 26th January, 2014

*Key words:*
OLS Estimator,
FGLS Estimators,
Combined Estimators,
Mean Square Error,
Linear Regression Model,
Monte Carlo Simulation.

ABSTRACT

The use of Ordinary Least Square (OLS) estimator for estimation of parameters of linear regression model in the presence of multicollinearity has been reported to produce imprecise estimates associated with large standard errors. This paper presents some combined estimators based on Feasible Generalized Linear Estimator (CORC and ML) and Principal Components (PCs) Estimator as alternative to multicollinearity estimation methods. A linear regression model with three uniformly distributed explanatory variables exhibiting high degree of multicollinearity ($\lambda \geq 0.7$) was considered through Monte Carlo studies. The experiments were conducted to assess and compare the performances of the various proposed combined estimators with their separate ones and the Ridge estimator using the Mean Square Error (MSE) criterion by ranking their performances at each parameter level and summing the ranks over the number of parameters. Results reveal that the proposed estimators of CORCPC1, MLPC1 and MLPC12 are generally better than the OLS estimator while CORCPC12 does the same at increased sample size. Furthermore, the combined estimator CORCPC1, recommended for usage, performs better than the Ridge estimator and it is either the best or does not perform too differently from the PC1 or PC12 estimator.



## INTRODUCTION

The Ordinary Least Squares (OLS) estimator has been known and reported to be Best Linear Unbiased Estimator (BLUE) of the classical linear regression model when all the fundamental assumptions of the model are non-violated (Fomby, 1984; Maddala, 2002). The use of the estimator for parameter estimation when the assumption of independence of explanatory variables is not valid (leading to multicollinearity) does not only produce imprecise estimates but also large standard errors. Consequently, insignificance of the true regression coefficient and misleading conclusions are often arrived at (Chatterjee and Hadi, 2006; Chatterjee, *et al.,* 2000). Various estimators including Ridge Regression Estimator (Hoerl, 1962; Hoerl and Kennard, 1970), estimator based on Principal Component Analysis Regression (Massy, 1965; Marquardt, 1970; Bock, *et al.,* 1973; Belsley *et al.,* 1980; Naes and Marten, 1988) and estimator based on Partial Least Squares (Helland, 1988; Helland, 1990; Phatak and Jony, 1997) have been developed to tackle this problem. Another problem associated with linear regression model is that of non-independence of error terms leading to autocorrelation. Using the OLS estimator for parameter estimation in the presence of autocorrelated error terms has been reported to yield inefficient but unbiased estimates, inefficient predicted values and underestimated sampling variance of the autocorrelated error terms (Johnston, 1984; Fomby, 1984; Chatterjee, *et al.,* 2000; Maddala, 2002). To compensate for the lost of efficiency, several feasible generalized least squares (FGLS) estimators including Cochrane and Orcutt (1949), Paris and Winstern (1954), Hildreth and Lu (1960), Durbin (1960), Theil (1971), the maximum likelihood and the maximum likelihood grid (Beach and Mackinnon, 1978) and Thornton (1982) have been developed. Consequently, this paper attempt to combine a method of handling multicollinearity (Principal Component Analysis) and that of autocorrelation together with the motivation of examining the performance of the resulting estimators (called combined estimators) in handling multicollinearity problem when there is no autocorrelation in the model.

***Corresponding author:** **Kayode Ayinde,** Department of Statistics, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Oyo State, Nigeria*

## MATERIALS AND METHODS

Consider the linear regression model of the form:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + U_t \qquad (1)$$

$$U_t \sim N(0, \sigma^2), \; t = 1, 2, 3...n$$

For Monte-Carlo simulation study, the parameters of equation (1) were specified and fixed as $\beta_0 = 4$, $\beta_1 = 2.5$, $\beta_2 = 1.8$ and $\beta_3 = 0.6$. The levels of multicollinearity among the independent variables were sixteen (16) and specified as: $\lambda(x_{11}) = \lambda(x_{12}) = \lambda(x_{13}) =$ 0.7, 0.8, 0.9, 0.95 and 0.99.

Furthermore, the experiment was replicated in 1000 times (R =1000) under four (4) levels of sample sizes (n =10, 20, 30, 50). The correlated uniform regressors were generated by using the equations provided by Ayinde (2007) and Ayinde and Adegboye (2010) to generate normally distributed random variables with specified intercorrelation. With P= 3, the equations give:

$$X_1 = \mu_1 + \sigma_1 Z_1$$
$$X_2 = \mu_2 + \rho_{12}\,\sigma_2 Z_1 + \sqrt{m_{22}}\,Z_2 \qquad (2)$$
$$X_3 = \mu_3 + \rho_{13}\,\sigma_3 Z_1 + \frac{m_{23}}{\sqrt{m_{22}}} Z_2 + \sqrt{n_{33}}\,Z_3$$

Where $m_{22} = \sigma_2^2(1-\rho_{12}^2)$, $m_{23} = \sigma_2\sigma_3(\rho_{23}-\rho_{12}\rho_{13})$ and $n_{33} = m_{33} - \frac{m_{23}^2}{m_{22}}$;

and $Z_i \sim N(0, 1)$ i = 1, 2, 3. In the study, we assumed $X_i \sim N(0, 1)$, i = 1, 2, 3. We further utilized the properties of random variables that cumulative distribution function of Normal distribution produces U (0, 1) without affecting the correlation among the variables (Schumann, 2009) to generate $X_i \sim U(0,1)\ i = 1,2,3$.

Having simulated the data, the technique adopted for the development of the combined estimator is very much similar to that of the Principal Component Estimator when used to solve multicollinearity problem.

Just like the Principal Component does its estimation using the OLS estimator by regressing the extracted components (PCs) on the standardized dependent variable, the combined estimators use the FGLS estimators, Cochrane and Orchutt (CORC) estimator (1949) and the Maximum Likelihood (ML) estimator (Beach and Mackinnon, 1978), by regressing the extracted components (PCs) on the standardized dependent variable. Unlike the OLS estimator which results back into the OLS estimator when all the PCs are used, advantageously, since the FGLS estimators require an iterative methodology for its estimation, the proposed combined estimators may not result back into the FGLS feasible estimators when all the possible PCs are used for the estimation. Consequently, the parameters of (2) are estimated by the following twelve (12) estimators: OLS, PC1, PC12, CORC, CORCPC1, CORCPC12, CORCPC123, ML, MLPC1, MLPC12, MLPC123 and Ridge as suggested by Scolve (1973) and described in Amemiya (1985). The Ridge estimator is an empirical Bayesian estimator. The prior is that coefficients are zero with a variance estimated from the data as the sums of squared of the fitted values of the dependent variable divided by the trace of the design matrix. The Ridge parameter in this case is a consistent estimate of the residual variance divided by the variance of the coefficient prior.

## RESULTS AND DISCUSSION

The mean square errors of $\beta_0$ of the estimators are graphically presented in Figure 1.

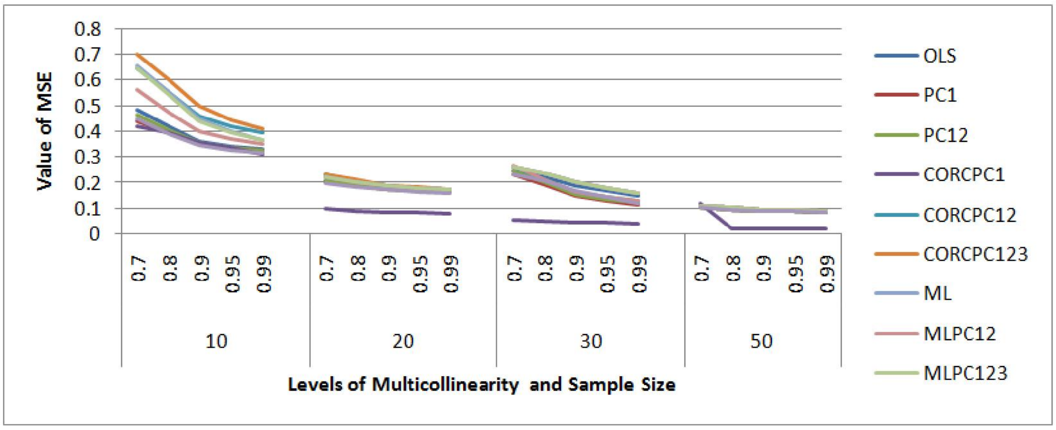


**Figure 1. Graphical Representation of $\beta_0$ Mean Square Error of the estimators at various levels of multicollinearity and sample size**

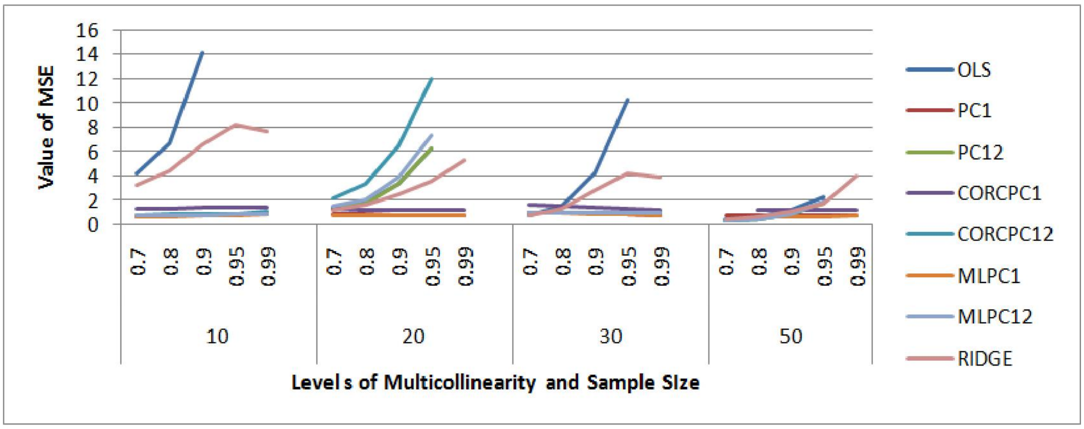


**Figure 2. Graphical Representation of $\beta_1$ Mean Square Error of the estimators at various levels of multicollinearity and sample size**

The figure does not capture that of MLPC12 and CORC estimators because the mean square error of the former is generally inefficient while that of the CORC estimator is grossly inefficient when the sample size is very small, n=10.

From Figure 1, it can be seen that the mean square error reduces as sample increases and that at each level of sample size the mean square error of the estimators reduces as multicollinearity level increases. The CORCPC1 estimator is generally most efficient estimator in estimating $\beta_0$. Figure 2 shows the graphical representation of the performances of the estimators on the basis of the mean square error of $\beta_1$ having removed the estimates of the CORC, ML, CORCPC123, MLPC123 estimators and other inefficient estimates. From Figure 2, it can be observed that at each level of sample size the mean square error of the OLS, CORCPC12 and MLPC12 estimators increases as multicollinearity level increases. The PC1 estimator is generally best while PC12, MLPC1 and CORCPC1 estimators compete very favorably.

In Figure 3 where the mean square errors of those competing estimators of $\beta_2$ are graphically, it is observed that at each level of sample size the mean square error of the OLS, PC12, MLPC12, RIDGE increases as the sample size increases. The PC1, CORCPC1 and MLPC1 estimators are generally efficient. Figure 4 shows the graphical representation of the competing estimators having removed the inefficient estimates. It is observed that at each level of sample size the mean square error of the OLS, PC12 and RIDGE estimators increases as the sample size increases. It further shows that the PC1, CORCPC1 and MLPC1 estimators are generally efficient. The summary of the performances of the estimator in term of their total rank over the model parameters at various levels of multicollinearity and sample size is given in Table 1. A sample of the Mean Square errors of the estimators that were ranked when n=20 is provided in the appendix. From the results in Table 1, it can be seen that the PC1 and PC12 estimator, and the proposed combined estimators, CORCPC1, MLPC1, MLPC12 and occasionally CORCPC12 estimators perform better than the OLS estimator.

**Table 1. Total rank of the Mean Square Error of the Estimators over the Parameters at different levels of multicollinearity and sample size**

| Sample size (n) | Estimators | Levels of Multicollinearity | | | | |
|---|---|---|---|---|---|---|
| | | 0.7 | 0.8 | 0.9 | 0.95 | 0.99 |
| | OLS | 28 | 28 | 27 | 26 | 27 |
| | PC1 | 6 | 7 | 6 | 8 | 7 |
| | PC12 | 14 | 15 | 14 | 16 | 16 |
| | CORC | 48 | 48 | 48 | 46 | 45 |
| | CORCPC1 | 11 | 11 | 13 | 12 | 11 |
| 10 | CORCPC12 | 29 | 30 | 30 | 30 | 30 |
| | CORCPC123 | 43 | 43 | 43 | 43 | 42 |
| | ML | 38 | 38 | 38 | 40 | 40 |
| | MLPC1 | 19 | 18 | 19 | 20 | 19 |
| | MLPC12 | 21 | 21 | 21 | 21 | 22 |
| | MLPC123 | 33 | 33 | 34 | 34 | 36 |
| | RIDGE | 22 | 20 | 19 | 16 | 17 |
| | OLS | 25 | 26 | 27 | 27 | 27 |
| | PC1 | 10 | 10 | 9 | 8 | 8 |
| | PC12 | 15 | 16 | 15 | 17 | 19 |
| | CORC | 47 | 47 | 47 | 47 | 47 |
| | CORCPC1 | 10 | 8 | 8 | 8 | 8 |
| 20 | CORCPC12 | 31 | 30 | 31 | 29 | 32 |
| | CORCPC123 | 43 | 43 | 43 | 43 | 42 |
| | ML | 36 | 36 | 36 | 37 | 37 |
| | MLPC1 | 22 | 22 | 19 | 18 | 18 |
| | MLPC12 | 21 | 21 | 23 | 23 | 25 |
| | MLPC123 | 34 | 34 | 35 | 35 | 35 |
| | RIDGE | 18 | 19 | 19 | 20 | 14 |
| | OLS | 14 | 30 | 31 | 31 | 31 |
| | PC1 | 26 | 10 | 7 | 7 | 8 |
| | PC12 | 24 | 14 | 15 | 16 | 17 |
| | CORC | 33 | 47 | 47 | 47 | 47 |
| | CORCPC1 | 15 | 11 | 11 | 11 | 11 |
| 30 | CORCPC12 | 40 | 22 | 23 | 25 | 25 |
| | CORCPC123 | 29 | 43 | 43 | 43 | 43 |
| | ML | 24 | 38 | 38 | 39 | 39 |
| | MLPC1 | 38 | 23 | 18 | 18 | 18 |
| | MLPC12 | 38 | 20 | 19 | 21 | 21 |
| | MLPC123 | 22 | 36 | 36 | 35 | 35 |
| | RIDGE | 9 | 18 | 24 | 19 | 17 |
| | OLS | 19 | 23 | 27 | 28 | 29 |
| | PC1 | 28 | 30 | 13 | 10 | 7 |
| | PC12 | 5 | 7 | 9 | 14 | 18 |
| | CORC | 35 | 39 | 43 | 44 | 42 |
| | CORCPC1 | 44 | 20 | 20 | 11 | 8 |
| 50 | CORCPC12 | 16 | 18 | 20 | 26 | 30 |
| | CORCPC123 | 33 | 37 | 41 | 42 | 42 |
| | ML | 29 | 33 | 37 | 38 | 38 |
| | MLPC1 | 48 | 40 | 25 | 18 | 19 |
| | MLPC12 | 12 | 14 | 18 | 23 | 24 |
| | MLPC123 | 29 | 33 | 37 | 37 | 38 |
| | RIDGE | 14 | 18 | 22 | 21 | 17 |

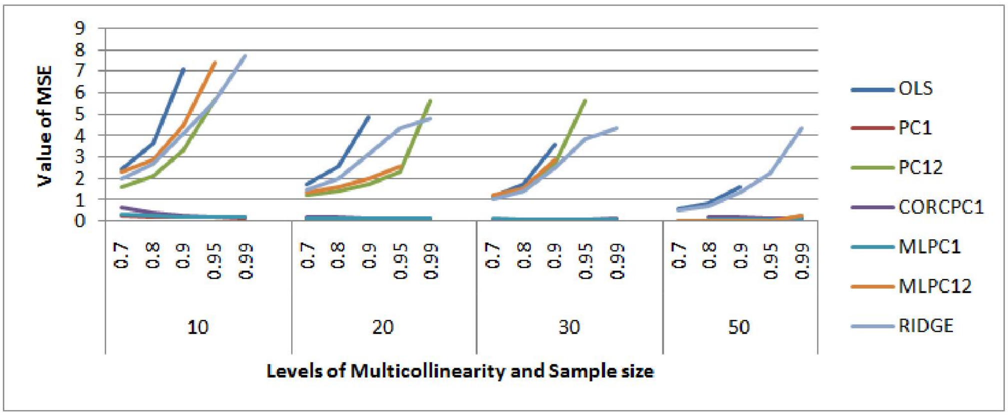


**Figure 3. Graphical Representation of $\beta_2$ Mean Square Error of the estimators at various levels of multicollinearity and sample size**

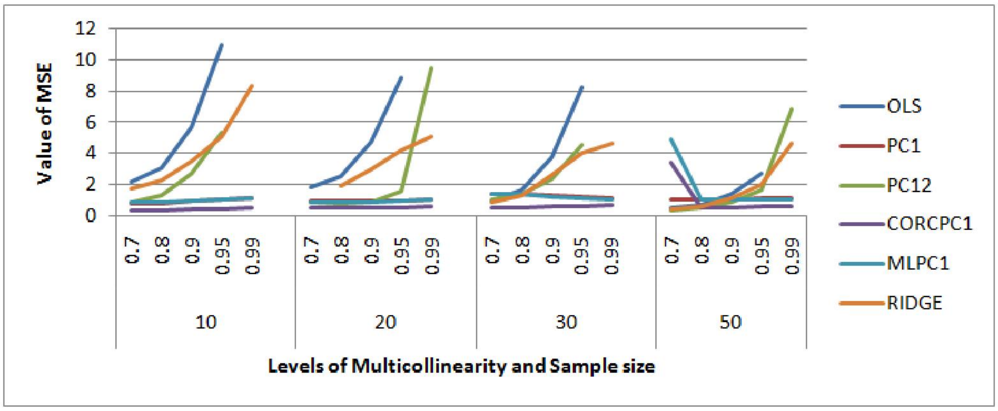


**Figure 4. Graphical Representation of $\beta_3$ Mean Square Error of the estimators at various levels of multicollinearity and sample size**

Moreover, the PC1 and CORCPC1 estimators perform better than the Ridge estimator even though the performance of the CORCPC12 estimator is not different from that of the Ridge. The best estimator is either PC1 or CORCPC1 and occasionally CORCP12.

### Conclusion

In this study, efforts have been made to combine two feasible Generalized Estimators with the estimator based on the principal components regression and compared their performances with that of the existing ones. These combined estimators when all the principal components are not used generally performed better than the OLS estimator and very precisely, the recommended combined CORCPC1 estimator is either best or performs not too differently from the best. This study has recommended some combined estimators as alternative to multicollinearity estimation methods.

## REFERENCES

Amemiya, Takeshi 1985. Advanced Econometrics, Harvard University Press, Cambridge, Mass.

Ayinde, K. 2007. Equations to generate normal variates with desired intercorrelation matrices. *International Journal of Statistics and System*, 2(2), 99 - 111.

Ayinde, K. and Adegboye, O.S. 2010. Equations for generating normally distributed random variables with specified intercorrelation. *Journal of Mathematical Sciences*, 21(2), 183 -203.

Beach, C. M. and Mackinnon, J.S. 1978. A Maximum Likelihood Procedure regression with autocorrelated errors. Econometrica, 46: 51 – 57.

Belsley, D.A., Kuh, E. and Welsch, R.E. 1980. Regression Diagnostics, Identifying Influence Data and Source of Collinearity. Wiley, New York.

Bock, M. E., Yancey, T. A. and Judge, G. G. 1973. Statistical consequences of preliminary test estimation in regression. *Journal of American Statistical Association*, 60, 234 -246.

Chatterjee, S. and Hadi, A.S. 2006. Regression by Example. 4th Edition, A Wiley- Interscience Publication, John Wiley and Sons.

Chatterjee, S.,Hadi, A.S. and Price, B. 2000. Regression by Example. 3rd Edition, A Wiley- Interscience Publication, John Wiley and Sons.

Cochrane, D. and Orcutt, G.H. 1949. Application of Least Square to relationship containing autocorrelated error terms. *Journal of American Statistical Association*, 44:32–61.

Durbin, J. 1960. Estimation of Parameters in Time series Regression Models. *Journal of Royal Statistical Society* B, 22:139 -153.

Fomby, T. B., Hill, R. C. and Johnson, S. R. 1984. Advance Econometric Methods. Springer-Verlag, New York, Berlin, Heidelberg, London, Paris, Tokyo.

Greene, W. H. 2003. Econometric Analysis. 5th Edition, Prentice Hall Saddle River, New Jersey 07458.

Helland I. S. 1988. On the structure of partial least squares regression. Communication is Statistics, Simulations and Computations, 17, 581 – 607.

Helland I. S. 1990. Partial least squares regression and statistical methods. *Scandinavian Journal of Statistics*, 17, 97 – 114.

Hildreth, C. and Lu, J.Y. 1960. Demand relationships with autocorrelated disturbances. Michigan State University. Agricultural Experiment Statistical Bulletin, 276 East Lansing, Michigan.

Hoerl, A. E. 1962. Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58, 54 – 59.

Hoerl, A. E. and Kennard, R. W. 1970. Ridge regression biased estimation for non-orthogonal problems. Technometrics, 8, 27 – 51.

Johnston, J. 1984. Econometric Methods. 3rd Edition, New York, Mc, Graw Hill.

Maddala, G. S. 2002. Introduction to Econometrics. 3rd Edition, John Willey and Sons Limited, England.

Marquardt, D. W. 1970. Generalized inverse, Ridge Regression, Biased Linear Estimation and Non – linear Estimation. Technometrics 12, 591 – 612.

Massy, W. F. 1965. Principal Component Regression in exploratory statistical research. *Journal of the American Statistical Association*, 60, 234 – 246.

Naes, T. and Marten, H. 1988. Principal Component Regression in NIR analysis: View points, Background Details Selection of components. *Journal of Chemometrics*, 2, 155 – 167.

Paris, S. J. and Winstein, C. B. 1954. Trend estimators and serial correlation. Unpublished Cowles Commision, Discussion Paper, Chicago.

Phatak, A. and Jony, S. D. 1997. The geometry of partial least squares. *Journal of Chemometrics*, 11, 311 – 338.

Schumann 2009. Generating correlated Uniform Variates. http://comisef.wikidot.com/tutorial:correlateduniformvariates.

Sclove, S.L. 1973. Least Squares Problems with Random Regression Coefficients. Technical Report No. 87, IMSS, Stanford University.

Theil, H. 1971. Principle of Econometrics. New York, John Willey and Sons.

Thornton, D. L. 1982. The appropriate autocorrelation transformation when autocorrelation process has a finite past. Federal Reserve Bank St. Louis, 82 – 102.


**APPENDIX:**

**Table 2. The Mean Square Error of the Estimators of the Parameters at different levels of multicollinearity when n = 20**

| Sample size (n) | Estimators | Levels of Multicollinearity | | | |
|---|---|---|---|---|---|
| | | MB0 | MB1 | MB2 | MB3 |
| 0.7 | OLS | 0.20554 | 1.40097 | 1.73222 | 1.79485 |
| | PC1 | 0.20277 | 0.88171 | 0.10643 | 0.93146 |
| | PC12 | 0.20706 | 1.28268 | 1.23922 | 0.84926 |
| | CORC | 0.25781 | 2.56129 | 2.50681 | 1.99219 |
| | CORCPC1 | 0.097641 | 1.36008 | 0.22663 | 0.49393 |
| | CORCPC12 | 0.23324 | 2.18664 | 1.94607 | 0.89466 |
| | CORCPC123 | 0.23388 | 2.42273 | 2.42999 | 1.97505 |
| | ML | 0.22269 | 1.67467 | 1.97394 | 1.89747 |
| | MLPC1 | 7.01814 | 0.88362 | 0.11291 | 0.93698 |
| | MLPC12 | 0.22187 | 1.50627 | 1.39684 | 0.86168 |
| | MLPC123 | 0.22291 | 1.66485 | 1.9699 | 1.89452 |
| | RIDGE | 0.19946 | 1.22403 | 1.46275 | 1.50425 |
| 0.8 | OLS | 0.18983 | 1.92185 | 2.53133 | 2.54163 |
| | PC1 | 0.18782 | 0.84869 | 0.098916 | 0.91659 |
| | PC12 | 0.19037 | 1.77917 | 1.42722 | 0.78608 |
| | CORC | 0.23045 | 3.74919 | 3.747 | 2.77029 |
| | CORCPC1 | 0.092439 | 1.31131 | 0.20687 | 0.48911 |
| | CORCPC12 | 0.20872 | 3.31923 | 2.39273 | 0.88064 |
| | CORCPC123 | 0.21047 | 3.52957 | 3.61933 | 2.75124 |
| | ML | 0.20389 | 2.3059 | 2.85082 | 2.68019 |
| | MLPC1 | 7.47266 | 0.85116 | 0.10548 | 0.91894 |
| | MLPC12 | 0.20216 | 2.10874 | 1.6123 | 0.80989 |
| | MLPC123 | 0.20418 | 2.29126 | 2.84549 | 2.67604 |
| | RIDGE | 0.18447 | 1.59596 | 1.98155 | 1.97937 |
| 0.9 | OLS | 0.17601 | 3.39758 | 4.88517 | 4.69199 |
| | PC1 | 0.17459 | 0.82241 | 0.093593 | 0.94198 |
| | PC12 | 0.17561 | 3.30882 | 1.78753 | 0.92484 |
| | CORC | 0.20381 | 7.05734 | 7.45986 | 5.00928 |
| | CORCPC1 | 0.086527 | 1.27618 | 0.19501 | 0.50536 |
| | CORCPC12 | 0.18769 | 6.4946 | 3.1438 | 1.26824 |
| | CORCPC123 | 0.18984 | 6.57107 | 7.15891 | 4.98619 |
| | ML | 0.18738 | 4.07326 | 5.43046 | 4.93338 |
| | MLPC1 | 8.0691 | 0.82688 | 0.099808 | 0.94161 |
| | MLPC12 | 0.18545 | 3.90694 | 2.02026 | 0.99919 |
| | MLPC123 | 0.18769 | 4.04533 | 5.42064 | 4.92596 |
| | RIDGE | 0.17121 | 2.4659 | 3.12479 | 3.01956 |
| 0.95 | OLS | 0.16963 | 6.27218 | 9.55116 | 8.85054 |
| | PC1 | 0.16839 | 0.81464 | 0.091654 | 0.98552 |
| | PC12 | 0.1689 | 6.26109 | 2.31526 | 1.61221 |
| | CORC | 0.19052 | 13.22224 | 14.79694 | 9.35538 |
| | CORCPC1 | 0.082969 | 1.26738 | 0.19341 | 0.52984 |
| | CORCPC12 | 0.17914 | 11.95322 | 4.02923 | 2.58509 |
| | CORCPC123 | 0.18075 | 12.27551 | 14.15545 | 9.32508 |

| | | | | | |
|---|---|---|---|---|---|
| | ML | 0.1798 | 7.48927 | 10.54155 | 9.29233 |
| | MLPC1 | 8.49181 | 0.82101 | 0.09743 | 0.98454 |
| | MLPC12 | 0.17818 | 7.321 | 2.6176 | 1.81164 |
| | MLPC123 | 0.18012 | 7.43768 | 10.52327 | 9.27881 |
| | RIDGE | 0.16505 | 3.61207 | 4.37342 | 4.18455 |
| | OLS | 0.16413 | 29.18772 | 46.70833 | 41.18924 |
| | PC1 | 0.16293 | 0.81372 | 0.091046 | 1.06605 |
| | PC12 | 0.16331 | 28.75434 | 5.66693 | 9.47457 |
| | CORC | 0.17887 | 59.93150 | 72.28310 | 43.30394 |
| | CORCPC1 | 0.078819 | 1.26846 | 0.19605 | 0.58000 |
| 0.99 | CORCPC12 | 0.17356 | 49.48396 | 9.30086 | 15.69635 |
| | CORCPC123 | 0.17423 | 55.79362 | 69.18285 | 43.19335 |
| | ML | 0.17334 | 34.54049 | 51.2222 | 43.21194 |
| | MLPC1 | 9.04656 | 0.82131 | 0.096142 | 1.06575 |
| | MLPC12 | 0.17220 | 33.03012 | 6.41319 | 10.81322 |
| | MLPC123 | 0.17367 | 34.31173 | 51.13714 | 43.15072 |
| | RIDGE | 0.15972 | 5.27550 | 4.77257 | 5.05291 |

*******