

Remedying *Multicollinearity* in Quantitative Analysis: A Simulation Studies

^{*1}Adamu Abubakar, ^{*2}U. F. Abbas, ^{*3}K. E. Lasisi

^{*1&2}Federal Polytechnic, Bauchi

^{*3}Abubakar Tafawa Balewa University, Bauchi

ABSTRACT

In multiple regression *modeling*, if there is a relationship between x_1 and x_2 , which are all the explanatory variables of y , then the relationship between x_1 and y or x_2 and y is not linear. Linear regression *modeling* is prone the problem of *Multicollinearity* as a result of violation of the assumption linearity. But, sometimes in real life data, it is difficult to satisfy the independence of predictor assumption due to the nature of the data leading to the problem of *Multicollinearity*. It causes interpretation misconceptions if the data set is contaminated with collinear variables. In this study, we restrict in the famous four methods which are Multiple Regression, Ridge Regression, Stepwise Regression and Partial Least Squares Regression. The aim of this study is to improve the remedial procedures for solving *Multicollinearity* problem in linear regression modeling. Findings from the simulation study shows that the partial least square method is the effective method when there is a change in sample size and number of predictors; it is capable to representing the data well when *Multicollinearity* problem is present.

ARTICLE INFO

Article History

Received: June, 2022

Received in revised form: July, 2022

Accepted: November, 2022

Published online: December, 2022

KEYWORDS

Multicollinearity, Impact, Remedy, Simulation, Parameter Estimate

INTRODUCTION

Concept of *Multicollinearity*

Multicollinearity is a condition of deficient data, which frequently is encountered in observational studies in which the investigator does not interfere with the study (Lukman *et al.*, 2015). It is obvious (from literature reviewed) that *Multicollinearity* till date is recognized as a very serious problem in linear regression modeling which lead up in wrong inferences. It is an important problem in linear regression modeling that has received increased attention in recent times (Ayinde *et al.*, 2015; Lukman *et al.*, 2015; Ismail and Manjula, 2016; Olanrewaju *et al.*, 2017).

The linear independence of the predictor variables must be met and is non-negotiable. If x_1 and x_2 are predictors of y and are related to some extent, the severity of such explanation will lead them not to explain the dependent variable.

There is a literature controversy among the researchers as to which technique is optimum when dealing with *Multicollinearity* in regression modeling. Wondola (2020) proposed partial least squares as the best method to deal with *Multicollinearity* problem. In his work he uses five (5) independent

variables and small sample size to test his assertion, but the question is, what will happen when the numbers of independent variables are much and when the sample size is large? Which method will produce efficient estimate when there is change in sample size and also change in the number of independent variables? His work is extended to improve the remedial procedures for solving *Multicollinearity* problem.

Source of *Multicollinearity*

There are several sources of *Multicollinearity* and a few of them are as follows:

- i. Use of defective data collection technique or employing unqualified personnel.
- ii. Faulty experimental design; violation of the pre-set protocol from the design of experiment or getting the wrong data.
- iii. Using a variable that is a sub-set of another variable within the same model which served as a duplicate or identical sister variable.
- iv. Error committed in terms of data coding; and so on. Rupa, 2022)

Impact of Multicollinearity

When multicollinearity exists in the data set, it is likely to:

- The estimates will have a large variances and covariance.
- Estimators (OLS) and their standard errors can be sensitive to small changes in the data, i.e. when add or delete an independent variable, the regression coefficients changes sensationally.
- There may be a strong and non-significant relationship, which is not possible.
- Confidence intervals on the regression coefficients may vary too wide with the parameter they estimate (over fit the data set).
- It can change the signs of the parameter estimated, a positive regression coefficient will be found to have a negative. (Rupa, 2022)

Remedying or Fixing Multicollinearity

Once Multicollinearity is identified in the data set, there are several ways of remedying it so as to create an accurate regression model, some of which are:

- Applying a Central Limit Theory by obtaining more data.
- Removing a collinear variable(s) (Kavish – Data Science Blogathon, 2022) (Stepwise Automated Regression).
- Use biased regression techniques in your model (employing biased regression methods).
- Do nothing: (use the traditional OLS method like multiple regression).
- Modify the weakest operator to a new look. (Rupa, 2022)

METHODOLOGY

The Ordinary Least Square (OLS)

The ordinary least squares (OLS) estimator is unbiased estimator. It is a tool that allow researcher to examine the relationship and significant effect of a number of independent variables on the dependent variable (Nael, 1994). The main goal of regression model is to minimize the error term and to estimate the parameters. The general linear regression model is given by:

$$y = x\beta + e \quad 3.1$$

Where Y is an $(n \times 1)$ response vector, X is an $(n \times p)$ matrix of n observations of p predictors.

$$E(e) = 0 \quad 3.2$$

And

$$v(e) = \delta^2 \quad 3.3$$

The solution turns out to be a matrix equation, defined by:

$$\beta = (X'X)^{-1}X'Y \dots\dots 3.4$$

In the Presence of Multicollinearity, (OLS) estimator could produce large variance, which leads to poor prediction (Ogee *et al.*, 2020). Some researchers are of the view that no single solution exists that will eliminate Multicollinearity totally (Allison, 2012; Cohen *et al.*, 2014; Alcacer *et al.*, 2018). Some opined that do nothing, live with what you have (Griffiths *et al.*, 1985; Green, 2003; Cohen *et al.*, 2014; Alcacer *et al.*, 2018).

Ridge Regression

The Ridge regression procedure is intended to overcome the ill-conditioned situation, and is used to improve the estimation of regression coefficients when regressors are correlated and the criterion of goodness of estimation considered in the ridge regression is the minimum total MSE (David, 2004). The ridge regression coefficients are given by:

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y \dots\dots 3.5$$

Where k is the ridge parameters and I is the identity matrix. (Imdad and Aslam, 2018a & b; Al-Hassan, 2007).

Stepwise Regression

Automatic variable selection methods have gained popularity in epidemiology and other fields for a number of reasons, including computational simplicity, relative ease of interpretation, selection of non-related independent variables (Multicollinearity) and their implementation (Rothman *et al.*, 2008). Stepwise algorithms still remain most widely used variable selection methods with collinear variable (Hutmacher & Kowalski, 2014).

Partial Least Square

Partial least square regression is a recent technique that generalizes and combines features from principal component analysis and multiple regression (Abdi, 2003; Ake & Ulf, 2016; Chadalavada, 2018). It decomposes both X and Y as a product of a common set of orthogonal factors and a set of

specific loadings, so, the independent variables are decomposed as $X = TP^T$ with $T^T T = I$, with I being the identity matrix, which means the issue of *Multicollinearity* has been taking care of (Abdi, 2003; Yizhou, 2009; Chadalavada, 2018; Babar, 2020).

Simulation Study and Analyses

Data set was simulated (Monte Carlo Simulation) using pseudo-random numbers where the response variable is $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_L x_{Li} + \varepsilon_i$ with $\varepsilon_i \approx N(0, 1)$.

Table 1: Simulated Data

p	y	n
5	$y = x_1 + x_2 + \dots + x_5 + N(0,1)$	10, 20, 30, 50
10	$y = x_1 + x_2 + \dots + x_{10} + N(0,1)$	15, 30, 70, 100
20	$y = x_1 + x_2 + \dots + x_{20} + N(0,1)$	30, 70, 150, 200
30	$y = x_1 + x_2 + \dots + x_{30} + N(0,1)$	45, 80, 190, 300
50	$y = x_1 + x_2 + \dots + x_{50} + N(0,1)$	75, 150, 250, 500

Detecting Multicollinearity

- Tolerance
The i^{th} tolerance value is defined as $1 - R^2$ where R^2 is the coefficient of determination (multiple determinations) for regression of the i^{th} independent variables.
- Variance Inflation Factor (VIF):
$$VIF = \frac{1}{1 - R^2}$$

- VIF is more than 10, predictors were seriously correlated and need to be address before modeling (Rupa, 2021).
- Eigen Values
All Eigen values must all be ≥ 0 , otherwise, the impact of *multicollinearity* is severe.

Table 2: VIF - P = 5

P = 5			Eigen Value
	Tolerance	VIF	
x1	0.005	220.165	0.112
x2	0.031	31.831	0.037
x3	0.007	133.651	0.001
x4	0.010	101.098	0.001
x5	0.854	1.171	0.000

Table 3: VIF - P = 10

P = 10			Eigenvalue	Excluded Variables			
	Tolerance	VIF			Beta In	Collinearity Statistics	
						VIF	Minimum Tolerance
			4.909				
x7	0	2063.173	0.055	x1	. ^b	.	0
x8	0.001	1394.417	0.036	x2	. ^b	.	0
x9	0.007	134.807	0	x3	. ^b	.	0
x10	0.279	3.59	1.55E-05	x4	. ^b	.	0
				x5	. ^b	.	0
				x6	. ^b	.	0

Table 4: VIF - P = 20

P = 20			Eigen value		Excluded Variables	
	Tolerance	VIF			VIF	Minimum Tolerance
			16.686	x5	.	0
x1	0.534	1.871	0.405	x6	.	0
x2	0.254	3.944	0.322			
x3	0.301	3.327	0.281			
x4	0.366	2.732	0.249			
x7	0.005	197.101	0.226			
x8	0.005	197.593	0.194			
x9	0.556	1.799	0.162			
x10	0.003	387.023	0.121			
x11	0.014	68.991	0.106			
x12	0.515	1.941	0.075			
x13	0.506	1.976	0.061			
x14	0.427	2.342	0.046			
x15	0.329	3.043	0.039			
x16	0.448	2.232	0.015			
x17	0.525	1.906	0.009			
x18	0.357	2.804	0.001			
x19	0.231	4.337	0			
x20	0.319	3.132	0			

Table 5: VIF - P = 30

P = 30			Eigen value	Excluded Variables			
	Tolerance	VIF			Beta In	Collinearity Statistics	
			12.547			VIF	Minimum Tolerance
x1	0.763	1.311	0.264	x2	. ^b	.	0
x5	0.005	194.229	0.23	x3	. ^b	.	0
x10	0.885	1.13	0.213	x4	. ^b	.	0
x11	0.835	1.197	0.2	x6	. ^b	.	0
x12	0.767	1.303	0.156	x7	. ^b	.	0
x15	0.678	1.475	0.127	x8	. ^b	.	0
x20	0.839	1.192	0.105	x9	. ^b	.	0
x21	0.719	1.391	0.072	x13	. ^b	.	0
x24	0.869	1.151	0.071	x14	. ^b	.	0
x27	0.027	36.937	0.014	x16	. ^b	.	0
x28	0.007	136.818	0.001	x17	. ^b	.	0
x29	0.01	102.219	0.001	x18	. ^b	.	0
x30	0.854	1.171	0	x19	. ^b	.	0
				x22	. ^b	.	0
				x23	. ^b	.	0
				x25	. ^b	.	0
				x26	. ^b	.	0

Table 6: VIF - P = 50

P = 50			Eigen Values	P = 50			Eigen Values
	Tolerance	VIF	43.599		Tolerance	VIF	43.599
x1	0.286	3.501	0.489	x26	0.384	2.607	0.103
x2	0.33	3.03	0.442	x27	0.298	3.357	0.095
x3	0.424	2.359	0.417	x28	0.346	2.894	0.09
x4	0.308	3.246	0.405	x29	0.335	2.986	0.084
x5	0.002	610.229	0.373	x30	0.371	2.699	0.083
x6	0.01	102.979	0.348	x31	0.229	4.375	0.074
x7	0.003	383.115	0.324	x32	0.368	2.72	0.072
x8	0.005	192.795	0.3	x33	0.303	3.299	0.062
x9	0.346	2.89	0.275	x34	0.33	3.034	0.06
x10	0.251	3.983	0.264	x35	0.237	4.218	0.052
x11	0.307	3.256	0.258	x36	0.41	2.441	0.046
x12	0.257	3.897	0.242	x37	0.258	3.883	0.045
x13	0.316	3.168	0.228	x38	0.36	2.777	0.036
x14	0.3	3.333	0.219	x39	0.323	3.093	0.034
x15	0.365	2.74	0.206	x40	0.387	2.584	0.03
x16	0.397	2.519	0.193	x41	0.323	3.097	0.025
x17	0.307	3.261	0.191	x42	0.388	2.577	0.022
x18	0.271	3.692	0.177	x43	0.291	3.431	0.016
x19	0.234	4.269	0.172	x44	0.308	3.246	0.015
x20	0.315	3.172	0.156	x45	0.281	3.555	0.013
x21	0.307	3.261	0.144	x46	0.36	2.778	0.01
x22	0.295	3.391	0.141	x47	0.291	3.439	0.001
x23	0.355	2.818	0.131	x48	0.404	2.477	0
x24	0.329	3.04	0.128	x49	0.276	3.623	0
x25	0.125	8.006	0.11	x50	0.458	2.181	8.07E-05

It is ensured that multicollinearity exists within these data sets corresponding to their respective variables highlighted **boldly** and the severity of such collinear variables need to be addressed before modelling.

The performance of such operators were ranked in order of strength and importance where the efficiency of the considered methods will be evaluated using R – Squared the coefficient of determination, mean squared error (MSE) and root mean square error (RMSE) to measure how well the regression model fit the data.

RESULT AND COMPARISON ANALYSES

Comparison Using R – Square Criteria (R^2)

(i) **Table 2:** P = 5: n = 10, 20, 30, 50

P = 5: n = 10, 20, 30, 50				
n	MR	RR	SR	PLS
10	0.7408	0.5116	0.5899	0.7431
20	0.5024	0.4913	0.4399	0.7196
30	0.3565	0.3487	0.3153	0.8115
50	0.5135	0.5092	0.5023	0.8022

(ii) **Table 3:** $P = 10: n = 15, 30, 80, 150$

$P = 10: n = 15, 30, 80, 150$				
n	MR	RR	SR	PLS
15	0.7340	0.2553	0.6104	0.7551
30	0.7340	0.2553	0.7340	0.7939
80	0.7325	0.2363	0.7358	0.8219
150	0.7312	0.2167	0.7315	0.8286

(iii) **Table 4:** $P = 20: n = 30, 80, 150, 200$

$P = 20: n = 30, 80, 150, 200$				
n	MR	RR	SR	PLS
30	0.6394	0.4878	0.7988	0.8770
80	0.2451	0.4701	0.3614	0.8189
150	0.2614	0.4236	0.4202	0.8601
200	0.2543	0.4483	0.4508	0.8880

(iv) **Table 5:** $P = 30: n = 70, 140, 220, 350$

$P = 30: n = 70, 140, 220, 350$				
n	MR	RR	SR	PLS
70	0.2533	0.2614	0.5327	0.7123
140	0.3620	0.2614	0.5477	0.7065
220	0.4474	0.2538	0.5151	0.8114
350	0.4770	0.2575	0.4440	0.8535

(v) **Table 6:** $P = 50: n = 75, 150, 300, 500$

$P = 50: n = 75, 150, 300, 500$				
n	MR	RR	SR	PLS
75	0.6551	0.2614	0.6954	0.7227
150	0.6707	0.2316	0.6423	0.7652
300	0.6450	0.2538	0.7020	0.8441
500	0.6575	0.2755	0.7201	0.8353

From among the four techniques, PLS out perform all other techniques in dealing with *Multicollinearity* problem in each case. It remained boldly with a change in the number of predictors (p) as well as change in the

sample size (n). On the other hand, RR technique is the most poorly performed method in dealing with *Multicollinearity* as it has the least values in each case.

Comparison Using Mean Square Error (MSE) Criteria

(i) **Table 7:** $P = 5: n = 10, 20, 30, 50$

$P = 5: n = 10, 20, 30, 50$				
n	MR	RR	SW	PLS
10	36807.0364	86681.2521	41594.1749	35018.7597
20	37411.4125	40977.8644	35088.9323	32500.1573
30	49362.3527	52041.8825	46894.0324	23618.7719
50	31071.0776	32060.6572	29803.2576	10227.8837

(ii) **Table 8:** $P = 10: n = 15, 30, 80, 150$

$P = 10: n = 15, 30, 80, 150$				
n	MR	RR	SW	PLS
15	24781.371	69376.083	30251.262	22865.069
30	19825.119	55500.858	19825.119	20610.909
80	18359.139	42743.826	17190.645	16871.932
150	17766.891	44274.093	17260.021	10061.895

(iii) **Table 9:** P = 20: n = 30, 80, 150, 200

P = 20: n = 30, 80, 150, 200				
n	MR	RR	SW	PLS
30	17771.5294	20712.9664	11801.6936	12964.3273
80	47565.8216	42462.4545	45027.6092	12010.6256
150	42696.3288	37453.8996	35476.8903	7939.8792
200	42106.5886	33936.9347	32512.4173	5774.7841

(iv) **Table 10:** P = 30: n = 70, 140, 220, 350

P = 30: n = 70, 140, 220, 350				
n	MR	RR	SW	PLS
70	35264.1827	35946.3399	30002.9166	10426.6563
140	46292.6207	31806.3327	37857.3292	10245.0835
220	41520.9903	30755.6541	31349.9249	8752.7251
350	42238.8403	34570.4111	31430.8222	6260.4491

(v) **Table 11:** P = 50: n = 75, 150, 300, 500

P = 50: n = 75, 150, 300, 500				
n	MR	RR	SW	PLS
75	44221.2112	60237.3575	29013.5012	13779.0035
150	42696.3288	26133.6323	24579.1845	10228.4905
300	40963.3717	23214.2400	18288.8839	8138.9267
500	41673.0171	21191.0616	16907.8269	6994.4787

RR has the largest MSE's in each case when compared with others whereas PLS has the lowest, indicating that PLS supersedes

other techniques when treating *Multicollinearity* while RR is the least.

Comparison Using Root Mean Square Error (RMSE) Criteria

(i) **Table 12:** P = 5: n = 10, 20, 30, 50

P = 5: n = 10, 20, 30, 50				
n	MR	RR	SR	PLS
10	191.8516	294.4168	203.9465	187.133
20	193.4203	202.4299	187.3204	180.278
30	222.1764	228.1269	216.5503	153.684
50	176.2699	179.0549	172.6362	101.133

(ii) **Table 13:** P = 10: n = 15, 30, 80, 150

P = 10: n = 15, 30, 80, 150				
n	MR	RR	SR	PLS
15	157.4210	263.3934	173.9289	151.212
30	140.8017	235.5862	140.8017	143.565
80	135.4959	206.7458	131.1131	129.892
150	133.2925	210.4141	131.3774	100.309

(iii) **Table 14:** P = 20: n = 30, 80, 150, 200

P = 20: n = 30, 80, 150, 200				
n	MR	RR	SR	PLS
30	133.3099	143.9200	108.6356	113.861
80	218.0959	206.0642	212.1971	109.593
150	206.6309	193.5301	188.3531	89.106
200	205.1989	184.2198	180.3120	75.992

(iv) **Table 15:** $P = 30$: $n = 70, 140, 220, 350$

$P = 30$: $n = 70, 140, 220, 350$				
n	MR	RR	SR	PLS
70	187.7876	189.5952	173.2135	102.111
140	215.1572	178.3433	194.5696	101.218
220	203.7670	175.3729	177.0591	93.556
350	205.5209	185.9312	177.2874	79.123

(v) **Table 16:** $P = 50$: $n = 75, 150, 300, 500$

$P = 50$: $n = 75, 150, 300, 500$				
n	MR	RR	SR	PLS
75	210.2884	245.4330	170.3335	117.384
150	206.6309	161.6590	156.7775	101.136
300	202.3941	152.3622	135.2364	90.216
500	204.1397	145.5715	130.0301	83.633

Result showed that among the techniques PLS outperform others in dealing with *Multicollinearity* problem as it has the lowest RMSE values in each case with changes in the number of predictors (p) and in the sample size (n). Whereas RR technique on the other hand is the most poorly performed method in dealing with *Multicollinearity* as it has the least values in each case hence need to be modified for a better performance.

CONCLUSION

It is sometimes difficult to satisfy the independence of predictor assumption of multiple regression due to the nature of the data leading to the problem of *Multicollinearity*. Despite the violation of such assumption, there are other techniques that handle data set that are contaminated with collinear variable(s). In this research, four alternative methods for solving *Multicollinearity* were comparatively evaluated against each other.

Findings showed that the best estimator for solving *Multicollinearity* problem, which has high efficiency, and performs well in capability, where it has minimum mean square error, the capability measure by root mean square error and coefficient of determination is Partial Least Squares (PLS), and more capability to resist its stance in dealing with *Multicollinearity* for both small and large samples and with small and large number of independent variables.

REFERENCE

Abdi, H. (2003). *Partial Least Squares (PLS) Regression*. The University of Texas at Dallas

Alcácer, J., Chung, W., Hawk, A., & Pacheco-de-Almeida, G. (2018) Applying random coefficient models to strategy research: Identifying and exploring firm heterogeneous effects. *Strategy Science*, 3(3): 533–553

Al-Hassan, Y.M. (2007) A Comparison Between Ridge Regression And Principle Component Regression Methods using Simulation Technique, AL al-Bayt University, 2007, (Unpublished)

Åke Björck and Ulf G. Indahl (2017) fast and stable partial least squares modeling: A benchmark study with theoretical comments; DOI: 10.1002/cem.2898

Allison, P. (2012). "When Can You Safely Ignore *Multicollinearity*?" *Statistical Horizons*. September 10, 2012. <http://statisticalhorizons.com/Multicollinearity>

Babar I, Ayed H, Chand S, Suhail M, Khan YA, Marzouki R (2021) Modified Liu estimators in the linear regression model: An application to Tobacco data. *PLoS ONE* 16(11): e0259991. <https://doi.org/10.1371/journal.pone.0259991>

Chadalavada, K. (2018) *Partial Least Squares Regression in Football Projections*; Senior Thesis in Mathematics.

Cohen, P., West, S. G., & Aiken, L. S. (2014) *applied multiple regression/correlation analysis for the behavioral sciences*. London: Psychology Press.

David A. W. (2004). *Ridge Regression as an Alternative to Ordinary Least*

- Squares: Improving Prediction Accuracy and the Interpretation of Beta Weights; Association for Institutional Research Number 92, summer, 2004; Enhancing knowledge. Expanding networks. Professional Development, Information Resources & Networking.
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Griffiths, W. E., Judge, G. G., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The theory and practice of econometrics*. New York: Wiley.
- Hutmacher MM, Kowalski KG. (2014) Covariate Selection in Pharmacometric Analyses: A Review of Methods. *Br J Clin Pharmacol*. 2014; 79(1):132–47.
- Imdad, M. U. and Aslam, M. (2018a) mctest: An R Package for Detection of Collinearity among Regressors. URL <https://cran.r-project.org/web/packages/mctest/>. R package version 1.2. [p328]
- Imdad, M. U. and Aslam, M. (2018b). Imridge: Linear Ridge Regression with Ridge Penalty and Ridge Statistics, URL <https://cran.r-project.org/package=Imridge>. R package version 1.2. [p326]
- Ismail, B; Manjula, S (2016) Estimation of linear regression model with correlated regressors in the presence of autocorrelation: *International Journal of Statistics and Applications* 6 (2), 35 – 39.
- Lukman, A. F., K. Ayinde, S. K. Sek, and E. Adewuyi. (2019). A modified new two-parameter estimator in a linear regression model. *Modeling and Simulation in Engineering* 2019:6342702. doi:10.1155/2019/6342702.
- Neale, M. C. (1994). *Mx: Statistical modeling* (2nd Ed). Department of Psychiatry: Medical College of Virginia.
- Ogee, A., Ellis, M., Scibilia, B. and Pammer, C. (2020) enough is enough! Handling *Multicollinearity* in Regression Analysis. The Minitab Blog.
- Olanrewaju, SO; Yahaya, HU; Nasiru, MO (2017). Effects of *Multicollinearity* and on some estimators in a system of regression equation. *European Journal of Statistics and Probability* 5 (3), 1 – 15.
- Rothman KJ, Greenland S, Lash TL. (2008) *Modern Epidemiology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008
- Rupa Singh (2022) *Multicollinearity* in Linear Regression: AI for Human Dignity | Thought Leader| Machine learning... | Founder and CEO at AI-Beehive | Building Published Jun 18, 2022
- Wondola, D.W. *et al* (2020) Partial Least Square (PLS) Method of Addressing *Multicollinearity* Problems in Multiple Linear Regressions (Case Studies: Cost of electricity bills and factors affecting it); IOP Conf. Series: Journal of Physics: Conf. Series **1463** (2020) 012006; IOP Publishing doi:10.1088/1742-6596/1463/1/012006
- Yizho Jiang (2009) Contributions to partial least squares regression and supervised principal component analysis modeling; Follow this and additional works at: https://digitalrepository.unm.edu/mat_h_etds