

## **Appraisal of the Budget Allocation Variables for Optimum Response**

**By**

<sup>1</sup>Umeh Maryrose Ngozi; <sup>2</sup>Ezeliora Chukwuemeka Daniel; <sup>1</sup>Mbeledeogu Njide N.; <sup>3</sup>Adinna B.O.

<sup>1</sup>*Department of Computer Science,* <sup>2</sup>*Department of Industrial and Production Engineering,* <sup>3</sup>*Department of Civil Engineering*  
<sup>1,2,3</sup>*Nnamdi Azikiwe University Awka, Anambra State, Nigeria*

**Abstract:** This work study the capital expenditure variables for optimization of the response. Partial least square method was used to optimize the variables of the capital expenditure.

### **Introduction to Partial Least Squares**

Partial least-squares (PLS) regression is a technique used with data that contain correlated predictor variables. This technique constructs new predictor variables, known as components, as linear combinations of the original predictor variables. PLS constructs these components while considering the observed response values, leading to a parsimonious model with reliable predictive power.

The technique is something of a cross between multiple linear regression and principal component analysis:

- Multiple linear regression finds a combination of the predictors that best fit a response.
- Principal component analysis finds combinations of the predictors with large variance, reducing correlations. The technique makes no use of response values.
- PLS finds combinations of the predictors that have a large covariance with the response values.

**The objective of the research** is to optimize the budget allocation variables using partial least square method

PLS therefore combines information about the variances of both the predictors and the responses, while also considering the correlations among them.

PLS shares characteristics with other regression and feature transformation techniques. It is similar to ridge regression in that it is used in situations with correlated predictors. It is similar to stepwise regression (or more general feature selection techniques) in that it can be used to select a smaller set of model terms. PLS differs from these methods, however, by transforming the original predictor space into the new component space.

The Statistics Toolbox™ function `pls regress` carries out PLS regression. [1]

### **Brief Explanation of Partial least Squares Regression**

Partial least squares regression (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models. Partial least squares Discriminant Analysis (PLS-DA) is a variant used when the Y is categorical.

PLS is used to find the fundamental relations between two matrices (X and Y), i.e. a latent variable approach to modeling the covariance structures in these two spaces. A PLS model will try to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. PLS regression is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values. By contrast, standard regression will fail in these cases.

The PLS algorithm is employed in PLS path modeling, [2][3] a method of modeling a "causal" network of latent variables (causes cannot be determined without experimental or quasi-experimental methods, but one typically bases a latent variable model on the prior theoretical

assumption that latent variables cause manifestations in their measured indicators). This technique is a form of structural equation modeling, distinguished from the classical method by being component-based rather than covariance-based. [4]

Partial least squares was introduced by the Swedish statistician Herman Wold, who then developed it with his son, Svante Wold. An alternative term for PLS (and more correct according to Svante Wold) is projection to latent structures, but the term partial least squares is still dominant in many areas. Although the original applications were in the social sciences, PLS regression is today most widely used in chemometrics and related areas. It is also used in bioinformatics, sensometrics, neuroscience and anthropology. In contrast, PLS path modeling is most often used in social sciences, econometrics, marketing and strategic management. [5]

## Underlying model

The general underlying model of multivariate PLS is a matrix of predictors and a matrix of responses and they are matrices respectively, projections of X (the X score, component or factor matrix) and projections of Y (the Y scores). The orthogonal loading matrices and the error terms assumed to be independent and they are identically distributed at random normal variables. The decompositions of X and Y are made so as to maximize the covariance of T and U.

## Algorithms of Partial Least Square

A number of variants of PLS exist for estimating the factor and loading matrices and most of them construct estimates of the linear regression between some PLS algorithms that is only appropriate for the case of a column vector, while others deal with the general case of a matrix. Algorithms also differ on whether they estimate the factor matrix as an orthogonal matrix or not. [6][7][8][9][10][11] The final prediction will be the same for all these varieties of PLS, but the components will differ. [12]

**The research method** applied is the use of the partial least square to predict the optimum variables.

**Table 1: Estimated Budget Allocation (in million naira)**

Educational(x <sub>2</sub> )	Health (x <sub>3</sub> )	Water Resources & Supply(x <sub>4</sub> )	Rural Electrification(x <sub>5</sub> )	Environmental Dev. & Mgt. (x <sub>6</sub> )	Finance & Investment (x <sub>7</sub> )	Commerce & Industry (x <sub>8</sub> )	Community Development (x <sub>9</sub> )	(Y)
3.7	20	17	5	2	20	14.1	3	104.7
2	20	14	5	2	20	9.5	2	99.5
2	20	14	5	2	20	9.5	2	99.5
5	29.9	20	5	3	20	22.4	5	131.6
18	23	42.5	12	4	4	38.4	8	177.2
20.5	30	65	20	4.1	17	49.6	9	307.9

**PLS Regression: C10 versus C1, C2, C3, C4, C5, C6, C7, C8, C9**

Method

Cross-validation None  
 Components to calculate Set  
 Number of components calculated 4

Analysis of Variance for C10

Source	DF	SS	MS	F	P
Regression	4	33094.0	8273.51	*	*
Residual Error	1	0.0	0.00		
Total	5	33094.0			

Model Selection and Validation for C10

Components	X Variance	Error	R-Sq
1	0.79725	2131.82	0.93558
2	0.92808	131.84	0.99602
3	0.99759	0.34	0.99999
4	1.00000	0.00	1.00000

Coefficients

	C10	standardized
Constant	-12.8482	0.000000
C1	0.8139	0.283093

C2	1.1439	0.118315
C3	0.9132	0.054899
C4	0.7386	0.188382
C5	2.7884	0.213122
C6	3.2040	0.039677
C7	1.5095	0.118757
C8	0.6305	0.128929
C9	2.2899	0.086143

Fits and Residuals for C10

Row	C10	Fits	Res	SRes
1	104.7	104.7	0.0000000	*
2	99.5	99.5	0.0000000	*
3	99.5	99.5	0.0000000	*
4	131.6	131.6	0.0000000	*
5	177.2	177.2	0.0000000	*
6	307.9	307.9	-0.0000000	*

Predicted Response for New Observations Using Model for C10

Row	Fit	SE Fit	95% CI	95% PI
1	104.7	0	(104.7, 104.7)	(104.7, 104.7)
2	99.5	0	(99.5, 99.5)	(99.5, 99.5)
3	99.5	0	(99.5, 99.5)	(99.5, 99.5)
4	131.6	0	(131.6, 131.6)	(131.6, 131.6)
5	177.2	0	(177.2, 177.2)	(177.2, 177.2)
6	307.9	0	(307.9, 307.9)	(307.9, 307.9)

X Residuals

Row	C1	C2	C3	C4	C5	C6
1	-0.0000000	-0.0000000	-0.0000000	-0.0000000	0.0000000	-0.0000000
2	0.0000000	-0.0000000	-0.0000000	-0.0000000	0.0000000	-0.0000000
3	0.0000000	-0.0000000	-0.0000000	-0.0000000	0.0000000	-0.0000000
4	-0.0000000	-0.0000000	-0.0000000	-0.0000000	0.0000000	-0.0000000
5	-0.0000000	0.0000000	-0.0000000	-0.0000000	-0.0000000	-0.0000000
6	0.0000000	0.0000000	-0.0000000	-0.0000000	-0.0000000	-0.0000000

Row	C7	C8	C9
1	0.0000000	-0.0000000	0.0000000
2	-0.0000000	-0.0000000	0.0000000
3	-0.0000000	-0.0000000	0.0000000
4	-0.0000000	-0.0000000	0.0000000
5	-0.0000000	-0.0000000	0.0000000
6	0.0000000	-0.0000000	-0.0000000

### Y Residuals

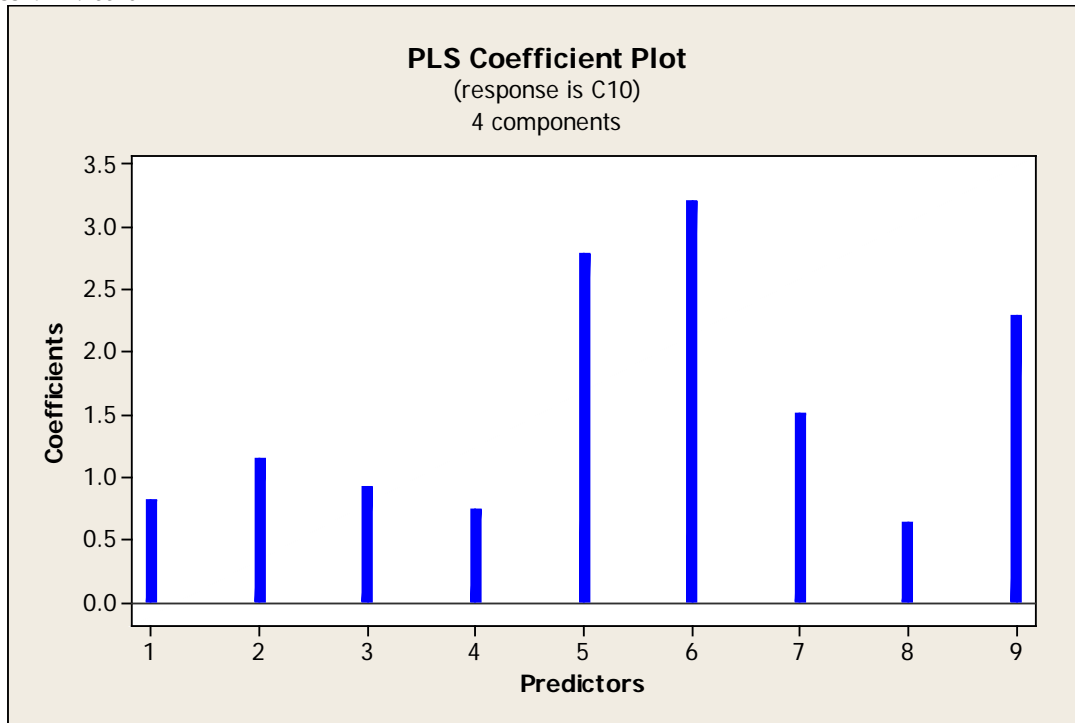
Row	C10
1	-0.0000000
2	-0.0000000
3	-0.0000000
4	-0.0000000
5	-0.0000000
6	-0.0000000

### X Calculated

Row	C1	C2	C3	C4	C5	C6	C7	C8	C9
1	19.9	3.7	20.0	17.0	5	2.0	20	14.1	3
2	25.0	2.0	20.0	14.0	5	2.0	20	9.5	2
3	25.0	2.0	20.0	14.0	5	2.0	20	9.5	2
4	21.3	5.0	29.9	20.0	5	3.0	20	22.4	5
5	27.3	18.0	23.0	42.5	12	4.0	4	38.4	8
6	92.7	20.5	30.0	65.0	20	4.1	17	49.6	9

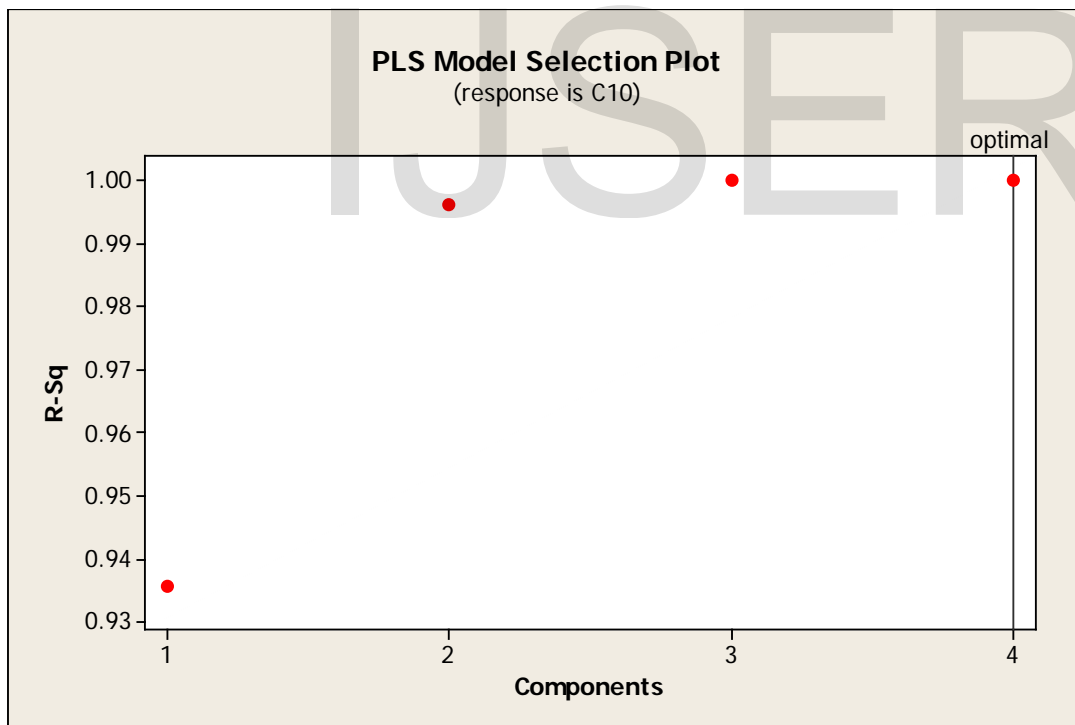
### Y Calculated

Row	C10
1	104.7
2	99.5
3	99.5
4	131.6
5	177.2
6	307.9



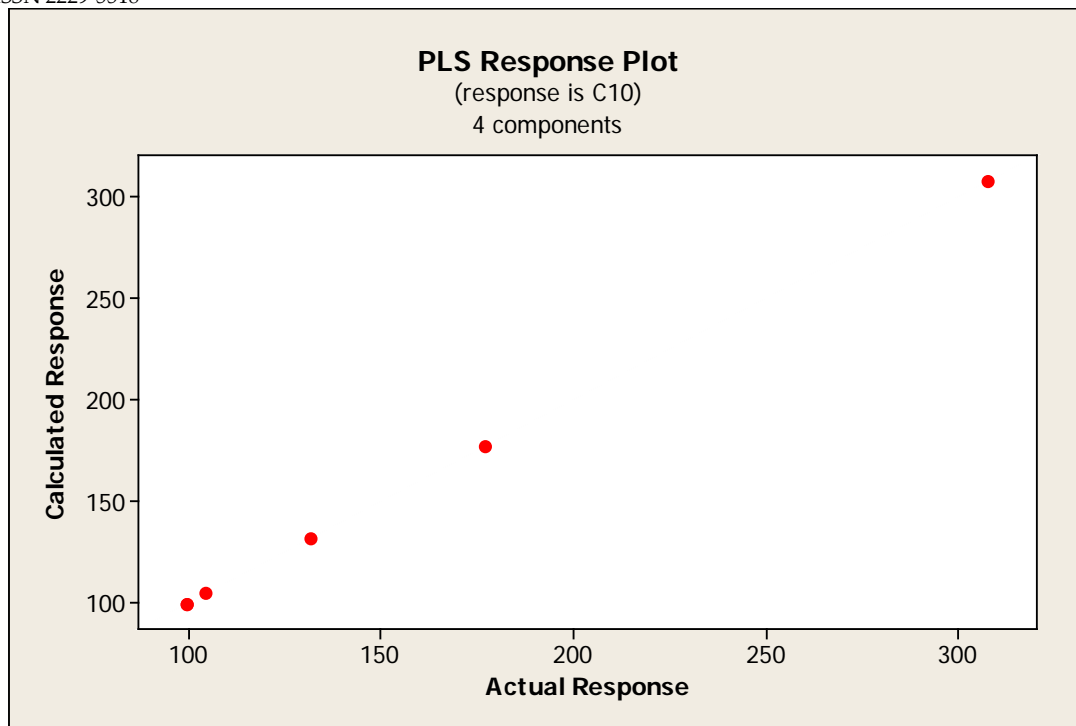
C10

PLS Coefficient Plot for



Plot for C10

PLS Model Selection



PLS Response Plot for

C10

**Table 2: The Coefficients and Predicts of the Model**

Coef1	Fit1	PFit1	PSEFit1	CLimit1	CLimit2	PLimit1	PLimit2
0.813869	104.7	104.7	0	104.7	104.7	104.7	104.7
1.143916	99.5	99.5	0	99.5	99.5	99.5	99.5
0.913191	99.5	99.5	0	99.5	99.5	99.5	99.5
0.738594	131.6	131.6	0	131.6	131.6	131.6	131.6
2.78836	177.2	177.2	0	177.2	177.2	177.2	177.2
3.204004	307.9	307.9	0	307.9	307.9	307.9	307.9
1.5095							
0.630518							
2.28991							

**Discussion and Conclusion:** This work studies the capital allocation variables. From the results, the model selection and validation show that at component 4, (component 4 means the variance of x values, the errors were reduced to zero and the R squared ( $R^2$ ) is 100%. This is the point that shows the optimum values of the capital allocation variables. Variance of X values (means the variation of the X value in four different ways) show also a result of 100%. At this point, the model is at the optimal point. In conclusion, the results of the data shows that the



model effective to optimize the variables and is necessary to also recommend it to the case study.

## References

1. <http://www.mathworks.com/help/stats/partial-least-squares.html>
2. Tenenhaus, M.; Esposito Vinzi, V.; Chatelinc, Y-M.; Lauro, C. (January 2005). "PLS path modeling" (PDF). *Computational Statistics & Data Analysis* 48 (1): 159–205. doi:10.1016/j.csda.2004.03.005.
3. Vinzi, V.; Chin, W.W.; Henseler, J. et al., eds. (2010). *Handbook of Partial Least Squares*. ISBN 978-3-540-32825-4.
4. Tenenhaus, M. (2008). "Component-based structural equation modelling" (PDF).
5. Wold, S; Sjöström, M.; Eriksson, L. (2001). "PLS-regression: a basic tool of chemometrics". *Chemometrics and Intelligent Laboratory Systems* 58 (2): 109–130. doi:10.1016/S0169-7439(01)00155-1.
6. Lindgren, F; Geladi, P; Wold, S (1993). "The kernel algorithm for PLS". *J. Chemometrics* 7: 45–59. doi:10.1002/cem.1180070104.
7. De Jong, S.; ter Braak, C.J.F. (1994). "Comments on the PLS kernel algorithm". *J. Chemometrics* 8 (2): 169–174. doi:10.1002/cem.1180080208.
8. Dayal, B.S.; MacGregor, J.F. (1997). "Improved PLS algorithms". *J. Chemometrics* 11 (1): 73–85. doi:10.1002/(SICI)1099-128X(199701)11:1<73::AID-CEM435>3.0.CO;2-#.
9. De Jong, S. (1993). "SIMPLS: an alternative approach to partial least squares regression". *Chemometrics and Intelligent Laboratory Systems* 18 (3): 251–263. doi:10.1016/0169-7439(93)85002-X.

10. Rannar, S.; Lindgren, F.; Geladi, P.; Wold, S. (1994). "A PLS Kernel Algorithm for Data Sets with Many Variables and Fewer Objects. Part 1: Theory and Algorithm". *J. Chemometrics* 8 (2): 111–125. doi:10.1002/cem.1180080204.
11. Abdi, H. (2010). "Partial least squares regression and projection on latent structure regression (PLS-Regression)". *Wiley Interdisciplinary Reviews: Computational Statistics* 2: 97–106. doi:10.1002/wics.51.
12. Trygg, J; Wold, S (2002). "Orthogonal Projections to Latent Structures". *Journal of Chemometrics* 16 (3): 119–128. doi:10.1002/cem.695.

IJSER