World Applied Sciences Journal 8 (4): 404-410, 2010
ISSN 1818-4952


# Dimension Reduction and Remedy of Multicollinearity Using Latent Variable Regression Methods

[1]*Shahla Ramzan and* [2]*Muhammad Inayat Khan*

[1]Department of Statistics, GC University, Faisalabad, Pakistan
[2]Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan

**Abstract:** Economic growth is the key goal of any economy. This study aims to examine the effect of some key macroeconomic determinants on Pakistan's economic growth. Two latent variable regression methods namely, Principal Component Regression and Partial Least Square Regression have been used for dimension reduction and to cope with the problem of multicollinearity. Time series data on major macroeconomic indicators from 1960 to 2007 has been used to determine the effect of these factors on the Economic Growth of Pakistan. The results show that education expenditures are negatively related to the growth, suggesting, the skill oriented education to be an important prerequisite to pick up the pace of growth. The empirical results also show that government consumption, openness of economy and private credit are important determinants of Pakistan's growth. The external debt is also negatively associated with growth, suggesting that relying on domestic resources is the best alternative to financial growth.

**Key words:** Multicollinearity · Dimension Reduction · Principal Component Regression (PCR) and Partial Least Square (PLS) Regression

## INTRODUCTION

Economic growth is the key goal of any economy. Macroeconomic variables; like inflation rate, savings, expenditures, investment and financial stability play an important role in determining the level of economic growth. Structural changes in the economy, natural calamities, political instabilities, global recessionary trends, self-feeding business cycles, etc. and the combined effect of all these factors is most commonly represented in the country's Gross Domestic Product (GDP).

A bulk of indicators affects the GDP of a country. These indicators are usually correlated with each other resulting in the problem of multicollinearity. The use of Ordinary Least Square (OLS) to estimate the parameters of the response function results in instability and variability of the regression coefficients. In this study Principal Component (PC) regression and Partial Least Square (PLS) regression are used for dimension reduction and to remedy multicollinearity and to explore the role of various factors on the economic growth of Pakistan after coping with the problem of multicollinearity.

In large-scale data mining and predictive modeling, especially for multivariate regression exercises, often one starts with a large number of possible explanatory/predictive variables. Therefore, variable selection and dimension reduction is a major task for multivariate statistical analysis, especially for multivariate regressions. A well known method in regression analysis for dimension reduction is called stepwise regression algorithm. One of the major limitations of the algorithm is that when several of the predictive variables are highly correlated, the tests of statistical significance that the stepwise method is based on are not sound, as independence is one of the primary assumptions of these tests. Often, many variables used as independent variables in a regression display a high degree of correlation, because those variables might be measuring the same characteristics. A high degree of correlation among the predictive variables increases the variance in estimates of the regression parameters. This problem is known as multicollinearity in regression literature. The parameter estimates in a regression equation may change with a slight change in data and hence are not stable for predicting the future.

**Corresponding Author:** Shahla Ramzan, Department of Statistics, G.C. University, Faisalabad, Pakistan.
E-mail: shahla_ramzan@yahoo.com.

A large theoretical and empirical literature exists relating a range of policy variables to economic growth in cross-country studies, but there have been few attempts to relate policy variables to growth in country-specific studies. Similarly, studies on determinants of economic growth in Pakistan have been very few [for example 1, 2, 3, 4].

The key macroeconomic determinants of economic growth are financial sector development, human sector development, openness of the economy, macro economic stability, etc. While studying these variables we observe that all the variables are related with one another, also they have a natural relation with respect to time, which is an indication of the correlation among these variables. The likely correlation that exists among the determinants of economic growth violates the basic assumption of multicollinearity. When we want to model these data, this high correlation handicaps our efforts. Therefore the remedial measures for multicollinearity are generally necessary.

This study is a contribution to fill these bottlenecks. The multicollinearity is removed using two latent variable regression methods namely, principal component (PC) regression and partial least square (PLS) regression. A Time series data set on major macroeconomic indicators from 1960 to 2007 has been used to determine the effect of these important factors on the Economic Growth of Pakistan.

## MATERIALS AND METHODS

Modeling the multiple time series data is an important task. Given the recent advent of unit root tests, the popularization and thorough derivation of the spurious regression problem [5] and the introduction of the concept of cointegration, there are now many more "rules and regulations" that the must be paid attention for properly modeling multiple time series.

A key feature of time series data that makes it more difficult to analyze than cross-sectional data is the fact that economic observations can rarely, if ever, be assumed to be independent across time. Most economic and other time series are related, often strongly related, to their recent past. While most econometric procedures can be used with both cross-sectional and time series data, more care is required while specifying econometric models for time series data before standard econometric methods can be justified.

Regression of one time series variable on one or more time series variables often can give nonsensical or spurious results. This phenomenon is known as spurious regression. One way to guard against it is to find out if the time series are cointegrated.

Cointegration means that despite being individually nonstationary, a linear combination of two or more time series can be stationary. Cointegration of two (or more) time series suggests that there is a long-run, or equilibrium, relationship between them.

Consider a regression model for two $I(1)$ variables, $Y_t$ and $X_t$, given by

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{1}$$

The term, $u_t$, is interpretable as the deviation from the relation in (1).
If $Y_t$ and $X_t$ cointegrate, then the deviation

$$u_t = Y_t - \beta_1 - \beta_2 X_t \tag{2}$$

is a stationary process with mean zero mean. In short, if the residuals obtained from regression of two or more time series variables are stationary or $I(0)$, the traditional regression methodology including the $t$ and $F$ tests is applicable to data involving nonstationary time series. The valuable contribution of the concepts of unit root, cointegration, etc. is to force us to find out if the regression residuals are stationary.

**Dimension Reduction and Multicollinearity:** When two or more independent variables in Multiple Regression are highly correlated, they both convey essentially the same information. In this case, neither may contribute significantly to the model after the other one is included. But together they contribute a lot. If both of the variables are removed from the model, the fit would be much worse. So the overall model fits the data well, but neither $X$ variable makes a significant contribution when it is added to the model. When this happens, the independent variables are collinear and the results show multicollinearity.

Different methods have been proposed which reduce the dimensions of the data and address the problem of multicollinearity. Adragni and Cook [6] and Chong and Jun [7] gave an overview of such regression methods which are simultaneously used for variable selection and heal for multicollinearity. Based on simulation results they highlighted the performance of these prediction methods. The PC regression and PLS regression techniques have been extensively used in different fields e.g Fekedulegn *et al.* [8] applied principal component regression to analyze the ecological data. Maitra and Yan [9] presented the procedures of principal component regression and PLS regression with an application to real world data. Muñiz *et al.* [10] used these two regression methods to predict the values of nutrients composition in milk. Mehvik and Wehrens [11] have described the procedure for implementing principal component regression and partial least square regression in R using PLS package. The mathematical foundations of the principal component analysis and partial least square algorithms have been presented by Glen *et al.* [12] and the partial least squares has been developed as a modification of principal component analysis. The PC regression and PLS regression have also been used in the field of agronomy. Ping *et al.* [13] compared these two methods for quantifying relationships between cotton yield or quality and soil properties.

**Principal Component Regression:** Principal component regression is a method for combating multicollinearity and results in estimation and prediction better than ordinary least squares. It is a widely used procedure for dealing with multicollinearity. In this method, the original m variables are transformed into a new set of orthogonal or uncorrelated variables called principal components of the correlation matrix. This transformation ranks the new orthogonal variables in order of their importance and the procedure then involves eliminating some of the principal components to effect a reduction in variance.

After elimination of the least important principal components, a multiple regression analysis of the response variable against the reduced set of principal components is performed using ordinary least squares estimation (OLS). Because the principal components are orthogonal, they are pair-wise independent and hence OLS method is appropriate. Once the regression coefficients for the reduced set of orthogonal variables have been calculated, they are mathematically transformed into a new set of coefficients that correspond to the original or initial correlated set of variables. These new coefficients are principal component estimators.

**Partial Least Squares Regression:** Partial least squares analysis is a multivariate statistical technique that allows comparison between multiple response variables and multiple explanatory variables. It is one of a number of covariance-based statistical methods which are often referred to as structural equation modeling (SEM). It was designed to deal with multiple regression when data has small sample, missing values, or multicollinearity. Haenlein and Kaplan [14] has provided a comprehensive overview of this technique.

Developed in the late 1960's by Herman O.A. Wold, partial least square regression was originally developed for use in the field of chemometrics. The PLS is a predictive technique which can handle many independent variables, even when these display multicollinearity. Partial least squares regression has been demonstrated on both real data and in simulations [15-18]. The success of PLS in chemometrics resulted in a lot of applications in other scientific areas including pharmacology, social sciences, physiology, bioinformatics and marketing research [18-22]. Zeng *et al.* [15] investigated the influence on generalization performance caused by the variation of the number of PLS components and the relationship between classification performance and regression quality of PLS on the training set.

Multiple Linear Regression (MLR) can be used when one has a large number of predictors. However, if the number of factors gets too large (for example, greater than the number of observations), it is likely to get a model that fits the sampled data perfectly but that will fail to predict new data well. This phenomenon is called over-fitting. In such cases, although there are many manifest factors, there may be only a few underlying or latent factors that account for most of the variation in the response. The general idea of PLS is to try to extract these latent factors, accounting for as much of the manifest factor variation as possible while modeling the responses well.

Assume $X$ is a $n{\times}p$ matrix and $Y$ is a $n{\times}q$ matrix. The PLS procedure extracts factors from both X and Y successively in such a way that covariance between the extracted factors is maximized. The PLS method can work with multivariate response variables.

PLS technique tries to find a linear decomposition of $X$ and $Y$ such that

$$X = TP^T + E$$

$$Y = UQ^T + F,$$

where

$T\ n{\times}g = X$-scores $\quad U\ n{\times}g = Y$-scores with $g \le p$

$P\ p{\times}g = X$-loadings $\quad Q\ 1{\times}g = Y$-loadings

$E\ n{\times}p = X$-residuals $\quad F\ n{\times}1 = Y$-residuals

The decomposition is finalized so as to maximize covariance between $T$ and $U$. Different algoritms are available to solve the PLS problem. However, all algorithms follow an iterative process to extract the $X$-scores and $Y$-scores. The factors or scores for $X$ and $Y$ are extracted successively and the number of factors extracted $(g)$ depends on the rank of $X$ and $Y$. In our case, $Y$ is a vector and all possible $X$ factors will be extracted.

Each extracted $X$-score is linear combinations of $X$. For example, the first extracted $x$-score $t$ of $X$ is of the form $t=Xw$, where $w$ is the eigenvector corresponding to the first eigenvalue of $X'YY'X$. Similarly the first $y$-score is $u=Yc$, where $c$ is the eigenvector corresponding to the first eigenvalue of $Y'XX'Y$. Note that $X'Y$ denotes the covariance of $X$ and $Y$.

Once the first factors have been extracted, the original values of $X$ and $Y$ are deflated as,

$$X_1 = X - tt'X$$

$$Y_1 = Y - tt'Y$$

The above process is repeated to extract the second PLS factors. The process continues until all possible latent factors $t$ and $u$ have been extracted, i.e., when $X$ is reduced to a null matrix. The number of latent factors extracted depends on the rank of $X$. This technique may also be used when there is a single response variable i.e., $Y$ is $n\times 1$ and $X$ is $n\times p$. A more detailed overview of partial least squares can be seen in [9, 23-25].

## RESULTS AND DISCUSSION

The variables included in the study are major macroeconomic indicators of economic growth described in Table 1. The data used for analysis have been collected from various sources [27, 28]. As the regression of one time series variable on one or more time series variables can produce spurious results therefore the data have been analyzed first as a time series. The degree of integration of each series used in the analysis has been determined with Augmented Dickey Fuller (ADF) test statistic, which indicates that all the variables are nonsignificant at 5% level. Thus all the variables involved are integrated of order one i.e. $I(1)$. Therefore the hypotheses that all the variables are stationary in the first difference, or integrated of order $I(1)$ cannot be rejected. So the series may be used to estimate co-integration regressions. The residuals obtained from OLS regression have been checked for the unit root using ADF test. The value of the ADF Test statistic is significant at 1% level of significance. Therefore the test rejects the null hypothesis that the residuals have a unit root. Hence the regression of these nonstationary time series is not spurious as the residuals are stationary.

Table 2 represents the OLS estimates of the regression coefficients, their standard errors and $t$-statistics. The overall model is highly significant with $R^2 = 99.8\%$ and adjusted $R^2 = 99.7\%$ therefore the regressors included in the model are 'important' for determining the economic growth. The value of Durbin-Watson statistic is 1.6488 suggesting that the residuals are serially uncorrelated. But the alarming point is that the individual $t$-ratios for some of the coefficients are nonsignificant. This is an indication of the presence of multicollinearity among the predictors. Looking at both the tolerance and VIF columns, only two of the independent variables (Financial depth and Openness) are causing less to multicollinearity, but the rest show a presence of high correlation.

Therefore to address this problem, principal component regression and partial least square regression techniques are used. Both techniques give solution to the problem of correlated regressors while reducing them into fewer dimensions. From the principal component analysis it was observed that the first principal component explains 78.8% of the total variation in the regressors. This is the most important variables as it itself contributing more than three-forth of the variation. First four principal components account for more than 98 percent of the variation in the response variable. The scree plot also showed that almost all the variation has been explained by the first 3 or 4 principal components. Therefore first four principal components are enough for estimating the regression model. Originally, 12 variables were explaining the almost the same amount of variation which is now explained by four variables only. Thus the dimension of 12 variables is now reduced to 4 variables with nearly the same amount of variation. Therefore only four principal components have been used to obtain the estimates of the coefficients of original variables. Using the relationship between PCR coefficients and the OLS coefficients, the estimated regression equation in terms of original variables is

$$\begin{aligned} Y = {} & 7.9479 + 0.1096(\text{PSE/LF}) + 0.0947\ (\text{MSE/LF}) + 0.0894\ (\text{HSE/LF}) + 0.0804\ (\text{OSE/LF}) \\ & - 0.0141(\text{ED/GDP}) + 0.1061\ (\text{M/GDP}) - 0.0136\ (\text{Openness}) + 0.0363(\text{Inflation}) \\ & + 0.0213\ (\text{GCE}) + 0.0213(\text{Investment}) + 0.0199(\text{CPS}) + 0.0204(\text{Edu}) \end{aligned}$$

These estimates of the regression coefficients show that all the predictors have positive influence to the economic growth. External Debt and Trade Openness has negative sign but are insignificant.

Table 1: Description of the Variables

| Variable | Description |
|---|---|
| Y | GDP per capita (real) |
| PSE/LF | Total Enrollment in Primary Schools as ratio of Labour Force (LF) |
| MSE/LF | Total Enrollment in Middle Schools as ratio of Labour Force (LF) |
| HSE/LF | Total Enrollment in High Schools as ratio of Labour Force (LF) |
| OSE/LF | Total Enrollment in Other Educational Institutions as ratio of Labour Force (LF) |
| ED/GDP | External Debt as percent of GDP |
| M/GDP | Financial Depth ($M_2$ as percent of GDP) |
| Openness | Openness of Trade = (Imports + Exports)/GDP |
| Inflation | Inflation |
| GCE | Govt. Consumption Expenditure |
| Investment | Investment |
| CPS | Credit to Private sector |
| Edu | Total Expenditures on Education |

Sources: 50 years of Pakistan (1947-1997)
Pakistan Economic Survey (various Issues)
Federal Bureau of Statistics
International Financial Statistics (IFS)
State Bank of Pakistan

Table 2: OLS Estimates of the Regression Coefficients and Multicollinearity Diagnostics

| Predictor | Coefficient | Standard Error | *t* | *p* value | $R^2$ | Tolerance | VIF |
|---|---|---|---|---|---|---|---|
| Constant | 7.7732 | 0.2698 | 28.81 | 0.00 | | | |
| PSE/LF | -0.0142 | 0.0609 | -0.23 | 0.82 | 0.986 | 0.014 | 70.174 |
| MSE/LF | 0.1712 | 0.0896 | 1.91 | 0.06 | 0.994 | 0.006 | 164.048 |
| HSE/LF | 0.0382 | 0.0877 | 0.44 | 0.67 | 0.994 | 0.006 | 175.226 |
| OSE/LF | 0.0602 | 0.0668 | 0.90 | 0.37 | 0.992 | 0.008 | 120.291 |
| ED/GDP | -0.0627 | 0.0165 | -3.80 | 0.00 | 0.896 | 0.104 | 9.589 |
| M/GDP | -0.0703 | 0.0478 | -1.47 | 0.15 | 0.634 | 0.366 | 2.734 |
| Openness | 0.0118 | 0.0182 | 0.65 | 0.52 | 0.708 | 0.292 | 3.427 |
| Inflation | -0.1318 | 0.0488 | -2.70 | 0.01 | 0.997 | 0.003 | 372.858 |
| GCE | 0.0674 | 0.0343 | 1.96 | 0.06 | 0.998 | 0.002 | 517.577 |
| Investment | -0.0557 | 0.0281 | -1.98 | 0.06 | 0.997 | 0.003 | 359.599 |
| CPS | 0.2189 | 0.0329 | 6.64 | 0.00 | 0.998 | 0.002 | 541.855 |
| Edu | -0.0247 | 0.0245 | -1.01 | 0.32 | 0.997 | 0.003 | 292.482 |

Table 3: Estimated Coefficients using Partial Least Square Regression

| | Coefficient PLS Regression | |
|---|---|---|
| Variable | GDPPC | Standardized GDPPC |
| Constant | 8.3387 | 0.0000 |
| PSE/LF | 0.0270 | 0.0298 |
| MSE/LF | 0.1834 | 0.2103 |
| HSE/LF | -0.0378 | -0.0457 |
| OSE/LF | 0.0986 | 0.1297 |
| ED/GDP | -0.0581 | -0.0873 |
| M/GDP | -0.0619 | -0.0172 |
| Openness | 0.0136 | 0.0111 |
| Inflation | -0.0868 | -0.2756 |
| GCE | 0.0540 | 0.2869 |
| Investment | -0.0703 | -0.3810 |
| CPS | 0.2096 | 1.1879 |
| Edu | -0.0179 | -0.1002 |

Now PLS regression is used to estimate the model for the data. Unlike PCA factors, PLS factors have multiple algorithms available to extract them. These algorithms are all based on iterative calculations. The Leave-One-Out method [26] for cross validation has been applied and the number of components to retain in the model is decided on the basis of PRESS statistic. The minimum value of the PRESS is 0.0266, at $8^{th}$ component. The X-variance indicates the amount of variance in the predictors that is explained by the model. The total amount of X-variance accounted for by the model is also a way to decide the number of components. The closer the X-variance value is to 1, the better the components represent the original set of predictors. The value of X-variance for 8 component model is 0.9985 i.e. almost all the variation in the predictors has been explained by the model. The variation explained in the dependent variable is 99.56%.

So both these values indicate the model formulated by 8 PLS components fits the data well with a high predictive power. The coefficients for the original model using PLS regression are indicated in Table 3.

The coefficient of external debt appears in the model with negative sign indicating its negative role on growth. There is a need to increase the foreign exchange earnings, cut off non-development expenditures and capture the depreciation of the value of rupee. More facilities should be given to foreign investors so that the inflow of foreign investment could increase in the country. When the foreign exchange earnings increase, the debt burden will also decrease.

As hypothesized, increase in education expenditures improve the human capital of the country and lead to enhance productivity of labor input. Results of this study interestingly do not conform to the aforesaid hypothesis. It can be attributed to many reasons. Nonetheless, it may reflect the lack of skill oriented education (the accepted phenomenon to raise GDP) in Pakistan's education system. The possibility of suboptimal allocation of resources can not be ruled out either. Education and training in general and skills development in particular, not only play a vital role in individual, organisational and overall national economic growth but are integral part of Human Resource Development (HRD). The results suggest that more emphasis should be placed on market oriented approach in education. This requires the overhauling of public school system not only in terms of curriculum but also in teaching methods.

Another interesting result is the negative association of $M_2$ to GDP ratio with economic growth. Perhaps, the negative coefficient reflects the deficiency of economy in adopting or imitating the technology that trickles through trade. There could also be the reason of maximum dependence of domestic economy on foreign manufactured goods.

Inflation is negatively influencing the growth in the estimated model; it implies that one percentage point change in inflation decreases GDP per capita by 0.13 percentage points. This quantification has specific meanings regarding role of inflation with GDP per capita; that is, statistically significant inflation with lower coefficient depicts its facilitating role in determining economic growth. Low and stable inflation specifically provides favorable environment in the growth of GDP. The increasing inflation in Pakistan is adversely affecting the per capita growth. So there is a need to control the rising inflation. Similarly, the coefficient of private sector credit (an indicator of financial sector development) is robust with positive sign in the model. It implies that increasing private credit facilitates the quality of inputs that is considered an important role player in enhancing economic growth. All of the other variables like openness of trade and government consumption expenditures confirm to the theory and appear with the positive coefficient and are statistically significant.

This study helps not only to explore the role of various macroeconomic indicators in the growth of the country but it can also be used in policy making. There is a need to improve internal resources and produce goods domestically so that the dependency on external sources may be reduced. The increasing inflation in Pakistan is adversely affecting the per capita growth hence there is a need to arrest the phenomenon of rising inflation.

## REFERENCES


1. Iqbal, Z., 1994. Macroeconomic Effects of Adjustment Lending in Pakistan. The Pakistan Development Review, 33(4): 1011-1031.
2. Iqbal, Z., 1995. Constraints to the Economic Growth of Pakistan: A Three-Gap Approach. The Pakistan Development Review, 34(4): 1119-1133.
3. Khilji, N.M. and A. Mahmood, 1997. Military Expenditures and Economic Growth in Pakistan. The Pakistan Development Review, 36(4): 791-808.
4. Shabbir, T. and A. Mahmood, 1992. The Effects of Foreign Private Investment on Economic Growth in Pakistan. The Pakistan Development Review, 31(4): 831-841.
5. Phillips, P.C.B., 1986. Understanding Spurious Regression in Econometrics. J. Econometrics, 33: 311-40.
6. Adragni, K.P. and R.D. Cook, 2009. Sufficient dimension reduction and prediction in regression. Philosophical Transactions of the Royal Society A, 367: 4385-4405.
7. Chong, I.G. and C.H. Jun, 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems, 78: 103-112.
8. Fekedulegn, B.D., J.J. Colbert, R.R. Hicks and M.E. Schucker, 2002. Coping with multicolinearity: An example on application of Principal Component Regression in Dendroecology. Research Paper NE-721, United States Department of Agriculture.
9. Maitra, S. and J. yan, 2008. Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. Casualty Actuarial Society Discussion Paper Program, pp: 79-90.

10. Muñiz, R., A.M. Pérez, C. de la Torre, C.E. Carleos, N. Corral and J.A. Baro, 2009. Comparison of Principal Component Regression (PCR) and Partial Least Square (PLS) methods in prediction of raw milk composition by vis-nir spectrometry. Application to development of on-line sensors for fat, protein and lactose contents. In XIX IMEKO World Congress, Fundamental and Applied Metrology, September 6-11, 2009, Lisbon, Portugal.
11. Mevik, B.H. and R. Wehrens, 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. J. Statistical Software, 18(2): 1-24.
12. Glen, W.G., W.J. Dunn III and D.R. Scott, 1990. Principal component analysis and partial least square analysis. Tethedron Computer Methodol., 2(6): 349-376.
13. Ping, J.L., C.J. Green, K.F. Bronson, R.E. Zartman and A. Dobermann, 2004. Identification of relationships between cotton yield, quality and soil properties. Agronomy J., 96: 1588-1597.
14. Haenlein, M. and A.M. Kaplan, 2004. A beginner's guide to partial least squares analysis. Understanding Statistics, 3(4): 283-297.
15. Zeng, X.Q., Li, G.H. and G.F. Wu, 2007. On the number of partial least squares components in dimension reduction for tumor classification. In The 2nd BioDM Workshop on Data Mining for Biomedical Applications (BioDM 2007).
16. Boulesteix, A.L. and K. Strimmer, 2007. Partial Least Squares: A Versatile Tool for the Analysis of High-Dimensional Genomic Data. Briefings in Bioinformatics, 8: 32-44.
17. Amador-HernaÂndez, J., L.E. GarcõÂa-Ayuso, J.M. FernaÂndez-Romero and M.D. Luque de Castro, 2000. Partial least squares regression for problem solving in precious metal analysis by laser induced breakdown spectrometry. J. Analytical Atomic Spectrometry, 15: 587-593.
18. Molfetta, F.A., A.T. Bruni, F.P. Rosselli and A.B.F. da Silva, 2007. A partial least squares and principal component regression study of quinone compounds with trypanocidal activity. J. Structural Chem., 18: 49-57.
19. Hulland, J.S., 1999. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. Strategic Management J., 20: 195-204.
20. Lobaugh, N.J., R. West and A.R. McIntosh, 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. Psychophysiol., 38: 517-530.
21. Nguyen, D.V. and D.M. Rocke, 2002. Tumor classification by partial least squares using microarray gene expression data. Bioinformatics, 18: 39-50.
22. Ryan, M.J., R. Rayner, and A. Morrison, 1999. Diagnosing customer loyalty drivers: Partial least squares vs. regression. Marketing Res., pp: 19-26.
23. Garthwaite, P.H., 1994. An Interpretation of Partial Least Squares. J. American Statistical Association, 89(425): 122-135.
24. Rosipal, R. and N. Krämer, 2006. Overview and recent advances in partial least squares. In Subspace, Latent Structure and Feature Selection Techniques, pp: 34-51. Springer.
25. Geladi, P. and B.R. Kowalski, 1986. Partial Least Squares Regression: A Tutorial. Analytica Chimica Acta, 185: 1-17.
26. McLeod, G., K. Clelland, H. Tapp, E.K. Kemsley, R.H. Wilson, G. Poulter, D. Coombs and C.J. Hewitt, 2009. A comparison of variate pre-selection methods for use in partial least squares regression: A case study on NIR spectroscopy applied to monitoring beer fermentation. J. Food Engineering, 90: 300-30.
27. Federal Bureau of Statistics, 1998. 50 Years of Pakistan: Volume I-Summary. Islamabad: Government of Pakistan.
28. Pakistan Economic Survey (various issues). Government of Pakistan, Finance Division, Economic Advisor's Wing, Islamabad.