**UtilitasMathematica**
**ISSN 0315-3681 Volume 118, 2021**

# Application of Principal Component Analysis and Hierarchical Regression Model on Kenya Macroeconomic Indicators

**Morris Kateeti Mbaluka, Dennis K. Muriithi, Gladys G. Njoroge**
[1]Chuka University, Kenya, morriskateeti@gmail.com
[2]Chuka University, Kenya, kamuriithi2011@gmail.com
[3]Chuka University, Kenya, gg.njoroge@gmail.com

## Abstract

The aim of this paper was to apply Principal Component Analysis (PCA) and hierarchical regression model on Kenyan Macroeconomic variables. The study adopted a mixed research design (descriptive and correlational research designs). The 18 macroeconomic variables data were extracted from Kenya National Bureau of Statistics and World Bank for the period 1970 to 2019. The R software was utilized to conduct all the data analysis. Principal Component Analysis was used to reduce the dimensionality of the data, where the original data set matrix was reduced to Eigenvectors and Eigenvalues. A hierarchical regression model was fitted on the extracted components, and R2 was used to determine whether the components were a good fit for predicting economic growth. The results from the study showed that the first component explained 73.605 % of the overall Variance and was highly correlated with 15 original variables. Additionally, the second principal component described approximately 10.03% of the total Variance, while the two variables had a higher positive loading into it. About 6.22% of the overall variance was explained by the third component, which was highly correlated with only one of the original variables. The first, second, and third models had F statistics of 2385.689, 1208.99, and 920.737, respectively, and each with a p-value of 0.0001<5% was hence implying that the models were significant. The third model had the lowest mean square error of 17.296 hence described as the best predictive model. Since component 1 had the highest Variance explained, and model 1 had a lower p-value than other models, Principal component 1 was more reliable in explaining economic growth. Therefore, it was concluded that the macroeconomic variables associated with the monetary economy, the trade and openness of the economy with government activities, the consumption factor of the economy, and the investment factor of the economy predict economic growth in Kenya. The study recommends that PCA should be utilized when dealing with more than 15 variables, and hierarchical regression model building technique be used to determine the partial variance change among the independent variables in regression modeling.

**Keywords:** Eigenvalues, Eigenvectors, Economic Growth, Macroeconomic Indicators, Principal Component Analysis.

## I. Introduction

A. Principal Component Analysis

The Principal Component Analysis (PCA) is technique through the preparation and analysis of a completely functional method that involved generating orthogonal axes in a single set, which run in descending order and determine the major directions of the sample variabilities [1]. The PCA technique is an essential tool in helping the researchers minimize the number of probable variables by grouping them into factors known as principal components. Using PCA is important as it helps remove the redundancy that exists in some of these variables [2]. PCA, therefore, analyses the variables and determines the constructs each of them measures, and then group those that measure the same constructs together.

The PCA model simplifies large data sets by reducing their dimensionality with no information loss. The PCA creates variables that are uncorrelated which end up maximizing the variance successfully. Eigen values and eigen vectors are used in determining the variance explained by variables when dealing with PCA. According to [2], it is difficult to fit a regression model with more than 15 variables. Therefore, to be able to determine the effect of many variables, it is necessary first to apply PCA then later frit a regression or any other model to the PCs obtained [3]. While applying PCA, acre must be taken to avoid the problem of over and under extraction. The components retained should be a true representation of the original data set matrix since all inferences are made based on the components and linked to the original data set.

There exist literature with precise utilization of PCA in different fields of study. For example, a study by

[4] utilized PCA in the assessment of the oil prices affecting food prices worldwide. The study used PCA in the determination of the effect of the macroeconomic index on food prices. The study looked at macroeconomic factors such as crude oil prices, the consumer price index, food production index, and GDP from 1961 to 2005 around the world. A proportion of variance (the Kaiser Criterion) and Scree plots was used in the analysis to figure out how many common components should be used. According to the macroeconomic index, the correlation coefficient ranged from 0.36 for the consumer price index to 0.87 for global GDP. Conclusion: The food production index influenced the macroeconomic index more than any other factor. Nevertheless, the researchers discovered a link between the oil price index and the index of food production. Oil prices, on the other hand, had no measurable impact on food costs. Although the study was highly effective, parallel analysis, which is beneficial in determining how many components should be preserved, was not utilized. There while applying PCA, both parallel analysis and Kaiser criterion should be used in the extraction and retaining of components. In the long run, this would help in reducing the problem of over and under extraction.

Principal components need to be extracted through various procedures to help minimize the problem of under and over extraction. Research by [5] utilized PCA in the evaluation of the examination results for secondary school. The main objective of the study was to determine the principal components in terms of individual subjects which play a key role in the students'

performance. The results presented a high correlation among all the subjects and found the highest variance in the first component. English subject was found to be the principal component, and hence its role was considered the most significant in the individual student's examination performance. Catelli plots and Kaiser Scree plots were used in the analysis. The problem of excess and under extraction remains, however, because the parallel analysis was never used to establish the number of components that should be maintained in this study. Additionally, the application of PCA was not optimal to this study since there were only a few variables (less than 15) being studied. An increase in the number of variables would improve the application of PCA and hence, in the long run, come out with components with a high level of variance explained.

Researchers needs to be careful and utilize the right methods according to the data under study to extract and retain the PCs. To effectively use Keiser's criterion, [6] advices that a researcher needs a more than 250 observations sample size to effectively use Keiser's criterion and contain more than 0.6 average commonalities in order for them to retain all factors with an eigenvalue that exceeds 1. In another suggestion by [7] a more than 300 observations sample size calls for a scree plot as the most effective factor extraction procedure However, [8] compared research that used different principal components extraction approaches, such as the Keiser criterion, scree plots, and parallel analysis. Researchers found that using parallel analysis to figure out how many components to preserve was the most successful method. To use the principal components in inferences, rename them into new variables after collecting the principal components. It is necessary to employ a different model in order to describe the association between variables evaluated using PCA. Because of this, a hierarchical regression model is born.

B. Hierarchical Regression Model

Multiple linear regression is, in most cases, used in behavioral and social statistics data analysis. When conducting a multiple regression analysis, the main objective is always to determine the best predictor. The regression model is used to identify all the predictors that support a study. For statistical control and investigation of incremental validity, the hierarchical regression model is used to assess how factors above and beyond previously entered predictors contribute. In hierarchical regression, the predictor variables are entered into the analysis one at a time. The hierarchical regression model helps the researcher gain control via the utilization of the adjusted coefficient of determination, hence showing the effect of the predictor variables effectively. Under hierarchical regression, the order of entry of the predictor variables is based on a specific theory. The entry order of the variables' entry can also be determined by the researcher. According to [9], the researcher usually knows more than the computer hence the need to select the order instead of letting the computer select the order of the variables as utilized in stepwise regression modeling. Most useful for studying predictor variables that are either closely associated or uncorrelated is the hierarchical regression model. A lot of the time, it's utilized to examine the impact of a predictor variable on the outcome. At each stage of the analysis, the coefficient of determination is calculated to ensure quality control [9]. After each principal component is added into the model, calculating the coefficient of

determination at each stage helps account for variance increases.

The multiple regression model does not explain the effects of the predictor variable on the dependent variable as well as the hierarchical regression model does. For instance, [10] carried out research to investigate the effects of macroeconomic variables on Pakistan's GDP. The research used principal comment analysis in conjunction with a multiple regression model to arrive at its conclusions. Three elements were identified and kept out of a total of 17 macroeconomic variables. The extracted components were all fitted with a multiple regression model, and it was discovered that they all had an impact on GDP. However, each predictor variable's effect was not clearly shown in this study. As a result, a hierarchical regression model might be utilized to understand the relationship between the predictor variables better. Some studies have found a hierarchical regression model to enhance the precision of estimates as compared to other conventional analysis methods. For instance, a study by [11] applied a hierarchical regression model to investigate the multiple paternal occupational exposures and neuroblastoma on offspring. The main study looked at the connection between fathers' work history and their children's risk of developing neuroblastoma. There were 405 patients in the study and 302 healthy controls. The effects of each exposure were estimated using conventional maximum likelihood as well as hierarchical regression. Comparing hierarchical regression to traditional analysis, the overall precision was considerably improved. Using hierarchical regression, the researchers found that the conventional approach's shortcomings may be mitigated by accounting for associated exposures, making some estimates more accurate using prior knowledge, and improving estimate precision. According to the findings of De Roos's research, the hierarchical regression model performed well in comparison to the others.

This study applied PCA and a hierarchical regression model on Kenya macroeconomic indicators. We were guided by three specific objectives that are applying PCA to reduce the dimensionality of Kenya Macroeconomic data and classify them into principal components, fitting a hierarchical regression model on the extracted Principal components in relation to economic growth in Kenya, determining the best predictors of economic growth.

## II. LITERATURE REVIEW

Principal Component Analysis

Karl Pearson invented PCA in 1901 [12]. Currently, exploratory data analysis and developing predictive models both use principal component analysis. A decomposition of the covariance matrix's auto values is used to arrive at this result. A principal component analysis (PCA) is used to analyze the data (Factor scores) [12]. PCA's primary objective is to build a linear combination of the variables under investigation in order to explain the covariance and variance of a random vector composed of random variables. The principal components are linear combinations.

Consider a random vector of interest $X' = (X1, X_2,..., X_N)$ with a covariance matrix $\sum$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p \geq 0$

We have the linear combinations as follows;

$$Y_1 = a_1'X = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \quad (1)$$
$$Y_p = a_2'X = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p$$
$$\vdots$$
$$Y_p = a_p'X = a_{p11}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p$$

We have $Var(Y_i) = a'_i \sum a_i$ and $Cov(Y_i Y_k) = a'_i \sum a_k$ where $i, k = 1,2,3,\ldots,p$

The principal components $Y_1$, $Y_2$, . . . , $Y_p$)Should, therefore, capture as much information as possible.(1) Let $\sum$ be the covariance matrix with the eigenvalue eigenvector pairs $(\lambda_1, \ell_1), (\lambda_2, \ell_2), \ldots, (\lambda_p, \ell_p),$ and$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n \geq 0$, then the $i$th principal component is given by:

$$Y_1 = \ell_1'X = \ell_{11}X_1 + \ell_{12}X_2 + \ldots + \ell_{1p}X_p \quad (2)$$

for $i$ = 1,2,3, . . . , $p$

It's worth noting that the variance of the $i$th principal component equals the $i$th eigenvalue.

$Var(Y_i) = \ell'_i \sum \ell' = \lambda'$ and $Cov(Y_i Y_k) = \ell'_i \sum \ell_k = 0$ where$i$ = 1,2,3, . . . , $p$ and $i \neq k$

Linear combinations of random variables produce the primary components. They are uncorrelated and have variances equal to the eigenvalues of Σ (the covariance matrix); thus, there is no need to make any assumptions regarding multivariate normality distributional assumptions in their construction [13].

The kth principal component's share of total variance can be expressed as follows:

$$K^{th} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \quad (3)$$

Where $\lambda$(is the eigenvalue of the $k$th PC. The first k PCs can explain the majority of the variance in population covariance; hence k variables can take the place of the original p variables with only a small drop in precision [13].

Hierarchical Regression Model

Advanced regression models use hierarchical regression as a model-building technique. Building successive linear regression models with more predictors is a common statistical technique. Our interest here is whether the next model explains the dependent variable better than the previous model [14]. If the difference of $R$" between the previous model and the Next model is statistically significant; we can say the added variables in the next model explain the dependent variable above and beyond the variables in the previous model. One of the advantages of the hierarchical regression model is that it displays the degrees of freedom well in most statistical software. Therefore, the hierarchical regression output and the statistical significance displayed by the regression are statistically significant and correct. It becomes easier to choose the best predictor after the analysis since decisions of entry of the variables were made manually and from research instead of being made through arbitration [14].

Separate level 1 models are created for each of the J levels two units in two-level hierarchical models [14]. Consider a situation in which the outcome or dependent variable is continuous, and there is a single continuous predictor or covariate X at level 1.

The level 1 models are of the form:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(X_{ij} - \bar{X}\right) + \varepsilon_{ij} \qquad (4)$$

Where$Y_{ij}$,is the predictor variable measured on the ith level 1 unit nested within the jth level 2 unit, $B$-, is the intercept for the jth level 2 unit, $X_{ij}$, is the level 1 predictor or covariate, $X$A is the grand mean of $X$',$B$!, is the regression coefficient linked with level 1 predictor X for the jth level 2 unit, and Ɛit is the random error linked with the ith level 1 unit nested within the jth level 2 unit.

In the level 2 models, we consider these regression coefficients ($B$-, and $B$!,) as well as covariates at a level 2 and link them all. W is a continuous level 2 predictor or covariate, and level 2 models have the following form:

$$B_{0j} = \gamma_{00} + \gamma_{01}W_j + V_{oj}; B_{1j} = \gamma_{10} + \gamma_{11}W_j + V_{1j} \qquad (5)$$

where ($B$-, and $B$!,) are the slope and intercept for the jth level 2 unit, $\gamma$--and $\gamma$-! are the overall mean slope and intercept adjusted for W, respectively, $W$,is the level 2 predictor or covariate, $\gamma$!- and $\gamma$!! are the regression coefficients linked with the level 2 predictor W relative to the level 2 slopes and intercept, respectively and $V$.,and $V$!, are the random effects of the jth level 2 unit on the intercept and slope, respectively, adjusted for W [14].

Predictor W can be modeled in its original metric or in relation to its grand mean, depending on the situation (similar to how the level 1 predictor is modeled in level 1). The combined model is obtained by substituting the level 2 model into the level 1 model.

$$B_{0j} = \gamma_{00} + \gamma_{01}W_j + V_{10}\left(X_{ij} - \bar{X}\right) + \gamma_{11}W_j\left(X_{ij} - \bar{X}\ldots\right) + V_{0j} + V_{1j}\left(X_{ij} - \bar{X}\ldots\right) + \varepsilon_{ij} \qquad (6)$$

## III. METHODOLOGY

### A. Research Design

The study made use of a variety of research approaches, including descriptive and correlative research designs. Descriptive design was used to help explain the characteristics of the study variables. Researchers employed a correlational research strategy to see what kind of relationship naturally exists between the variables they were studying [15].

### B. Data Collection

This study aimed at examining 18 macroeconomic variables for the period between 1970 and 2019. This period was found effective since it covers over a 30-year period. The data set matrix was a $P \times N$matrix where m is the macroeconomic variables. Secondary data was gathered from the World Bank and the Kenya National Bureau of Statistics (KNBS) websites for this study. For analysis, the data was saved in excel sheets and then imported into statistical software.

C. Principal Component Analysis

In this study, the high dimensional data set containing macroeconomic variables were summarized by data matrices X with n columns and p rows, the rows representing the observations, and the columns the variables as shown below;

$$X = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pn} \end{bmatrix} (P \times N) \quad (7)$$

Where $a'$,; $i$ = 1,2, … , $n$ and $j$ = 1,2,3, … , $p$ are the macroeconomic variables

PCA can use the covariance matrix to extract PC values if the original multivariate data set is not available. The correlation matrix was utilized instead of the covariance matrix to calculate PC since distinct variables in the data set were measured using different units and hence had varied variances.

The major component is obtained by decomposing the random vector's covariance matrix. Using the covariance matrix relative to the converted vectors, we may calculate the components of the random vector after it is transformed. The principal components are derived from the covariance matrices of the originally standardized variables in this scenario. These are the same as the principal components extracted from the original variables using the correlation matrix.

The PCA model can be represented by;

$$\mu_{m\times 1} = W_{m\times d}\, X_{d\times 1} \quad (8)$$

Where μ, m-dimensional vector = projection of X - the original time-series data.

X = the original matrix

W = E transpose

d = dimensional data vector ($m$ << $d$).

The m projection vectors that maximize the Variance of u called the principal axes are given by the eigenvectors$e1$, $e2$, … , $e$m of the data set covariance matrix S, it corresponds to the m largest non-zero eigenvalues. $\lambda 1$, $\lambda 2$, … $\lambda$m.

The data set's covariance matrix S can be found from:

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x-\mu)(x-\mu)^T \quad (9)$$

Where μ is the mean vector of x.

The eigenvectors ei can be found by solving the set of equations:

$$(S - \lambda_i I)e_i = 0\,; \quad i = 1,2,\ldots,d \quad (10)$$

Where λi are the eigenvalues of S.

The Eigenvalues give the total Variance explained a given principal component. It is obtained by the sum of squared component loadings across all items for each component. Non-zero eigenvalues are good, while negative eigenvalues are not suitable since Variance cannot be

negative, negative eigenvalues. Eigenvalues close to zero indicate the presence of multicollinearity since all the Variance can be taken up by the first component. Since the communality in a PCA for a single item is 1, a component with an eigenvalue larger than one is usually treated as a factor or primary component. The Scree Plot, which depicts the eigenvalue (total Variance explained) by the individual components, was used to choose the components in this investigation. Additionally, the dimensionality of these vectors was visually reduced using a Bi-plot visualization technique. The weight of each eigenvalue was provided through eigenvectors. The parallel analysis aided in deciding how many major components should be kept.

The eigenvectors are calculated and then ranked based on the magnitude of the related eigenvalues, as shown below. This is followed by selecting the top-m eigenvectors. Next, the PCA projection matrix is computed as follows:

$$W = E^T \tag{11}$$

Where E has the same eigenvectors as its columns, here W is an $m \times d$ matrix.

PCA reduces dimensionality by finding the principal components with the maximum Variance in the input feature vector components without transforming the input space. This research looked at explanations with a probability of more than 75%. Orthogonal rotation methods assume that the factors in the analysis are uncorrelated. Marczyk [16] lists four different orthogonal methods: "equinox, orthomax, quartimax, and varimax." In contrast, oblique rotation methods assume that the factors are correlated. Since factors are believed to be uncorrelated, this study employed orthogonal rotation. Marczyk [16], recommended rotating with varimax (orthogonal) or Promax (oblique). Varimax rotation maximizes the disparities between high and low loadings on a given factor while minimizing the variations of the loadings within the factors. As a result, the rotated component matrix was constructed using the varimax technique in this work. The component loadings are obtained by multiplying the eigenvector with the square root of the eigenvalue. The obtained factor loadings are used to determine each item's correlation with its respective principal component. The sum of squared loadings across components is the communality. "Loadings of 0.30 or higher can be considered significant, or at least salient" [17]. Thus, variables with loadings of 0.30 or higher are grouped into more than one factor.

The matrix of estimated factor loadings, L, is given by;

$$L = \left[\sqrt{\widehat{\lambda_1}\widehat{\varrho_1}}, \sqrt{\widehat{\lambda_2}\widehat{\varrho_2}}, \dots, \sqrt{\widehat{\lambda_m}\widehat{\varrho_m}}\right] \tag{12}$$

Where L is the matrix of factor loadings and $\lambda'$ Is ith eigenvalue.

D. Hierarchical Regression Model

Using a hierarchical regression model, you can see how changing one variable affects another. The general regression model of the study was

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_1 X_i + \varepsilon \quad (13)$$

This was expanded as;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (14)$$

Where;
Y = Gross Domestic Product
β0 = constant
βi = regression coefficients i = 1,2,3,…,n Xi = the principal components
$\varepsilon$ = error term
The successive hierarchical models were as follows;

Model 1 $$Y = \beta_0 + \beta_1 X_1 \quad (15)$$

Model 2 $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (16)$$

Model i $$Y = \beta_0 + \beta_1 X_A + \beta_2 X_2 + \cdots + \beta_n X_n \quad (17)$$

Predictability changes linked with predictor variables (Principal Components) entered later in the analysis over and above those given by predictor variables entered earlier in the analysis are the primary focus of the hierarchical regression model. In this research, the change in R2 was computed by entering predictor variables into the analysis in different steps. The order of the predictor variables was based on the variations explained by each component. The component with the largest Variance was entered first, then others followed simultaneously with the component with the lowest Variance entered last. Change in R2, change in F, and change in p values were the statistics of greatest interest. The change in R2 helped in determining the variation explained by each of the principal components. "The focus on R2 rather than on the β or the structure of the coefficients had less attention given to how the predictor variables were reevaluated on the basis of their corresponding βs and structure of the coefficients when the predictors were added to the analysis".
The coefficient of determination gives the proportion of the variability in the dependent variable that is accounted for by each model. Usually, the coefficient varies between zero and 1($0 \leq R" \leq 1$). The component that explains the highest variation is the one with a coefficient of determination close to one.
Trustrum [18], noted that a well-fitted regression model usually results in the predicted values being close to the observed data values. If no informative independent variables exist, the mean model can be employed, which utilizes the mean for each projected value. The proposed regression model, on the other hand, should fit better than the mean model. In regression analysis, three statistics are commonly used to assess model fit: R-squared, overall F-test (ANOVA), and Root Mean Square Error (RMSE), all of which are based on the sum of squares.

The F-test was used since the study goal is explanatory and predictive in nature. The F-test was determined using an analysis of Variance. The F-test compares "the null hypothesis of all regression coefficients being equal to zero to the alternative hypothesis of at least one regression coefficient not being equal to zero. R-squared equals zero is an equivalent null hypothesis" [19]. The F-test determines whether the proposed relationship between the predicted variable and the predictors is statistically reliable. A significant F-test indicates that the observed R-squared is reliable, and the data does not give spurious results. The p-value for the F-test is compared to the significance level, say 0.05. If the p-value is found to be less than the significance level, then it is concluded that the regression model fits the data better than the mean model. Such findings indicate that the independent variables (principal components) in the model improve the fit.

## IV. RESULTS AND DISCUSSION

### A. Descriptive Statistics

Before the data was subjected to any specific analysis, descriptive analysis of the predictor variables and the tread analysis of the predicted variable were carried out to help explain the distribution and behavior of the variable.

TABLE I: DESCRIPTIVE STATISTICS

| Variables (Natural Logs) | Mean | Std. Dev. | Correlation of Variation | Skewness | Kurtosis |
|---|---|---|---|---|---|
| *lnGDP* | 6.884 | 1.762 | 0.255 | 0.633 | 2.022 |
| *lnAQPROD* | 25.674 | 2.094 | 0.081 | -0.093 | 1.754 |
| *lnBM* | 14.949 | 0.233 | 0.015 | -0.007 | 2.648 |
| *lnCERPROD* | 3.144 | 0.234 | 0.074 | 0.613 | 2.536 |
| *lnDCPS* | 21.826 | 0.855 | 0.039 | 0.081 | 2.203 |
| *lnEXGS* | 17.936 | 1.791 | 0.099 | 0.108 | 3.119 |
| *lnFDINIC* | 23.254 | 1.048 | 0.045 | 0.325 | 2.225 |
| *lnGGFCE* | 21.435 | 0.956 | 0.044 | 0.211 | 2.249 |
| *lnGCFC* | 21.656 | 1.004 | 0.046 | 0.440 | 2.247 |
| *lnGDS* | 21.082 | 0.707 | 0.033 | 0.078 | 2.293 |
| *lnGNE* | 23.323 | 1.087 | 0.046 | 0.331 | 2.164 |
| *lnHHSNPISHS* | 22.896 | 1.147 | 0.050 | 0.312 | 2.107 |
| *lnIMGS* | 22.087 | 1.010 | 0.045 | 0.241 | 2.072 |
| *lnINF* | 2.2671 | 0.667 | 0.294 | -0.429 | 3.626 |
| *lnLINTR* | 2.7396 | 0.358 | 0.131 | 0.566 | 2.809 |
| *lnECHR* | 3.4932 | 1.006 | 0.288 | -0.417 | 1.503 |
| *lnREM* | 18.848 | 1.673 | 0.088 | 0.051 | 1.876 |
| *lnPOPPTT* | 17.093 | 0.448 | 0.026 | -0.214 | 0.851 |
| *lnUNEMP* | 1.0222 | 0.024 | 0.024 | -0.571 | 3.314 |

Where;

lnGDP = Natural Log of Gross Domestic Product

lnBM = Natural Log of broad money (current LCU

lnDCPS = Natural Log of domestic credit to the private sector (% of GDP)

lnEXGS = Natural Log of exports of goods and services (current US$)

lnFDINIC = Natural Log of foreign direct investment, net inflows (BoP, current US$)
lnGGFCE = Natural Log of general government final consumption expenditure (current US$)
lnGCFC = Natural Log of gross capital formation (current US$)
lnGDS = Natural Log of gross domestic savings (current US$)
lnGNE = Natural Log of gross national expenditure (current US$)
lnHHSNPISHS = Natural Log of households and NPISHs Final consumption expenditure (current US$)
lnIMGS = Natural Log of imports of goods and services (current US$)
lnINF = Natural Log of inflation, consumer prices (annual %)
lnLINTR = Natural Log of lending interest rate (%)
lnREM= Natural Log of personal remittances received (current US$)
lnCERPROD= Natural Log of cereal production (metric tons)
lnAQPROD = Natural Log of aquaculture production (metric tons)
lnECHR = Natural Log of the official exchange rate (LCU per US$, period average)
lnPOPTT = Natural Log of population, total
lnUNEMP = Natural Log of unemployment, total (% of the total labour force)

Table I indicates that from 1970 to 2019, the mean Gross Domestic Product (GDP) was 23.254 per year, with a standard deviation of 1.048, indicating that GDP fluctuated from year to year. The mean of foreign direct investment for the same period was 17.936 every year, and a standard deviation of 1.791, while Inflation had a mean of 2.267 and a standard deviation of 0.667. Based on the results, imports of goods and services generated a mean of 22.087 each year and a standard deviation of 1.011. Considering the values of Kurtosis, which is a measure of a distribution's tail behavior, the distribution showed a platykurtic behavior because all the values as indicated in Table I were less than 3. Additionally, the data is considered to be normal if the Skewness values are between -3 and 3. As per the findings of this research, all Skewness values were between -3 and 3, showing that the data used in the analysis was normal. Finally, the study examined the coefficient of variation (CV), which is a statistical measure of the dispersion of data points in a data series around the mean. A coefficient of variation is good if it is less than 1, while if it is more than 1, it implies a relatively high variation (the standard deviation is greater than the mean value). The more precise the estimate, the lower the coefficient of variation. Because of this, the coefficient of variation can be used to compare the degree of variation between different data sets. As it can be seen in Table I, all of the coefficients of variation were less than 1, indicating that the data set had low variation.

B. Sampling Adequacy Test -Bartlett's Test of Sphericity

TABLE II: KMO AND BARTLETT'S TEST

| Parameter | Measure | Statistic | Remark |
|---|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | 0.865>0.6 | PCA recommended for analysis |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2470.514 | Significance |
| | Df | 153 | |
| | p-value | 0.0001<0.05 | |

1) Test Results

$\chi 2 = 2470.514$; df = 153; $p<0.0001$

2) Statistical Decision

The null hypothesis that the correlation matrix of the variables is not significantly different from an identity matrix was rejected since the p-value was 0.0001 was less than 0.05 level of significance. This implied that the correlation matrix of the variables is significantly different from an identity matrix, and thus the sample correlation matrix did not come from a population in which the correlation matrix is an identity matrix. If the KMO statistics exceed 0.6 and Bartlett's test of sphericity is statistically significant (p-value is less than 5%), then PCA is usually recommended for analysis. "A Kaiser-Meyer-Olkin measure of sampling adequacy that is greater than 0.7 is regarded to be a good sign that PCA is useful for the variables under consideration, according to the rule of thumb". From the results in Table II, the correlations matrix was appropriate for component analysis because the KMO statistics was 0.865, which was more than the recommended 0.7. The Bartlett's Test in Table II was precisely sufficient for the data under study. This is because Bartlett's Test of Sphericity was used to test the difference between the correlation matrix for variables and the identity matrix was 2470.514. This showed that there was a significant difference. As a result, the correlation matrix for the measured variables differed significantly from the identity matrix and so remained consistent with the matrix's factorable premise.

C. Communalities

The sum of the squares of the loadings of the iit variables on the 'n' common components, that is, the iit commonality, was carried out in the results presented in Table III.

TABLE III: COMMUNALITIES

| Variables (Natural Logs) | Initial | Extraction |
|---|---|---|
| *lnBM* | 1.000 | 0.982 |
| *lnDCPS* | 1.000 | 0.886 |
| *lnEXGS* | 1.000 | 0.979 |
| *lnFDINIC* | 1.000 | 0.664 |
| *lnGGFCE* | 1.000 | 0.975 |
| *lnGCFC* | 1.000 | 0.983 |
| *lnGDS* | 1.000 | 0.860 |
| *lnGNE* | 1.000 | 0.987 |
| *lnHHSNPISHS* | 1.000 | 0.985 |
| *lnIMGS* | 1.000 | 0.982 |
| *lnINF* | 1.000 | 0.888 |
| *lnLINTR* | 1.000 | 0.853 |
| *lnREM* | 1.000 | 0.918 |
| *lnCERPROD* | 1.000 | 0.685 |
| *lnAQPROD* | 1.000 | 0.848 |
| *lnECHR* | 1.000 | 0.932 |
| *lnPOPPTT* | 1.000 | 0.973 |
| *lnUNEMP* | 1.000 | 0.793 |

Extraction Method: Principal Component Analysis.

All the macroeconomic indicators ranging from Broad money to total unemployment had a similar pattern and were highly correlated since the communalities of each indicator were greater than 0.65.

Furthermore, all the variables highly influence economic growth under the study period, as indicated by the high correlation.

D. Scree Plot and Parallel Analysis

The scree plot and parallel analysis to help in the determination of the principal components to be retained were carried out, and the results are presented in Fig. 1.

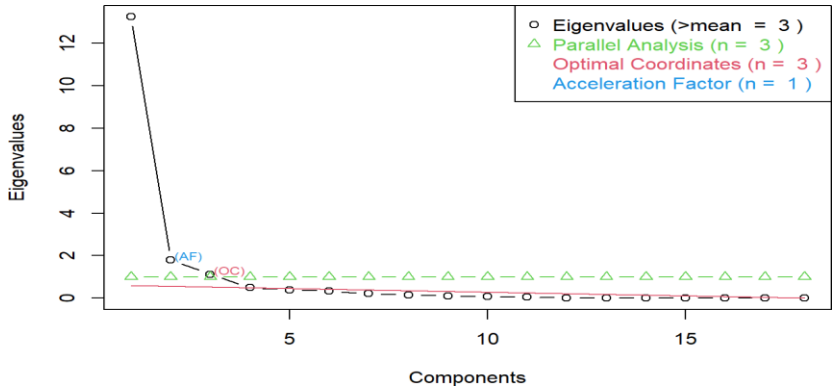


Fig.1. Scree plot.

To visually examine the Eigenvalues for inflection points, we used the Scree test. The visual representation looks to have a downward slope after the third PC, indicating that the three preceding PCs may be precisely summarized to be reflective of the variables in totality. As seen in the scree graphic, the use of Eigenvalues retrieved three principal components from the data. In addition, we used parallel analysis to avoid over-and under-extraction. After parallel analysis,

three components were chosen for further examination. The three components extracted and retained are a representation of all the original variables. Each of the three components contains a number of original variables. However, the number of original variables differs from one component to another. The number of variables in each component was identified using the rotated component matrix.

E. Total Variance Explained

The eigenvalues were used to carry out a test of the variation of the components, and the results are summarized in Table IV.

TABLE IV: TOTAL VARIANCE EXPLAINED

| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| Component | Total | % of Variance | Cumulative % | Total | % Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 13.249 | 73.605 | 73.605 | 13.249 | 73.605 | 73.605 | 12.994 | 72.190 | 72.190 |
| 2 | 1.806 | 10.034 | 83.639 | 1.806 | 10.034 | 83.639 | 1.957 | 10.870 | 83.060 |
| 3 | 1.119 | 6.217 | 89.856 | 1.119 | 6.217 | 89.856 | 1.223 | 6.796 | 89.856 |
| 4 | 0.493 | 2.738 | 92.594 | | | | | | |
| 5 | 0.390 | 2.169 | 94.763 | | | | | | |
| 6 | 0.334 | 1.855 | 96.618 | | | | | | |
| 7 | 0.208 | 1.158 | 97.776 | | | | | | |
| 8 | 0.151 | 0.840 | 98.616 | | | | | | |
| 9 | 0.106 | 0.587 | 99.203 | | | | | | |
| 10 | 0.073 | 0.403 | 99.606 | | | | | | |
| 11 | 0.043 | 0.238 | 99.843 | | | | | | |
| 12 | 0.016 | 0.089 | 99.932 | | | | | | |
| 13 | 0.007 | 0.037 | 99.969 | | | | | | |
| 14 | 0.003 | 0.019 | 99.988 | | | | | | |
| 15 | 0.001 | 0.007 | 99.995 | | | | | | |
| 16 | 0.001 | 0.004 | 99.999 | | | | | | |
| 17 | 0.000 | 0.001 | 100.000 | | | | | | |
| 18 | 6.172E-06 | 3.429E-05 | 100.000 | | | | | | |

From Table IV and considering the eigenvalue-one criterion, the first component explained 73.61 percent of the overall Variance while the second one explains about 10.03 percent of the total Variance. The 3rd, 4th, 5th, 6th, and 7th PCs explained around 6.217%, 2.74%, 2.17%, 1.86%, and 1.16 percent of the total variation, respectively. The remaining components described less than one percent of the total Variance each. Clearly, the first component explained the largest variation, and as we progress from one component to the next, a descending trend emerges. Additionally, both the first and second components cumulatively explained approximately 83.64 percent of the overall Variance. By employing the orthogonal Varimax technique to produce the

uncorrelated factor structures, the study found that the summarized total variation in the original set of data variables per the three retained components was about 89.86%.

F. Rotated Component Matrix

The rotating component matrix aids in the decision-making process. It includes the calculated correlations between each variable as well as the estimated main components. It comprises the calculated main components as well as the estimated correlations between the variables. As shown in Table V, the Rotated Component Matrix displays the loadings for each item on each rotating component.

TABLE V: ROTATED COMPONENT MATRIX

| Variables (Natural Logs) | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| *lnBM* | 0.944 | | |
| *lnDCPS* | 0.888 | | |
| *lnEXGS* | 0.983 | | |
| *lnFDINIC* | 0.797 | | |
| *lnGGFCE* | 0.983 | | |
| *lnGCFC* | 0.989 | | |
| *lnGDS* | 0.889 | | |
| *lnGNE* | 0.988 | | |
| *lnHHSNPISHS* | 0.985 | | |
| *lnIMGS* | 0.985 | | |
| *lnINF* | | | -0.927 |
| *lnLINTR* | | 0.871 | |
| *lnREM* | 0.939 | | |
| *lnCERPROD* | 0.818 | | |
| *lnAQPROD* | 0.919 | | |
| *lnECHR* | 0.831 | 0.489 | |
| *lnPOPPTT* | 0.940 | | |
| *lnUNEMP* | | 0.804 | 0.357 |

Table V indicates that component 1 was highly correlated with 15 original variables, which included; broad money, domestic credit to the private sector, gross capital formation, population total, exports of goods and services, general government final consumption expenditure, foreign direct investment-net inflows, personal remittances received, gross domestic savings, gross national expenditure, aquaculture production (metric tons), households and NPISHs Final consumption expenditure, imports of goods and services, imports of goods and services, cereal production (metric tons), and official exchange rate. This component mostly resembles the monetary economy, the trade and openness of the economy with government activities, the consumption factor of the economy, and the investment factor of the economy. Likewise, the lending interest rate and total unemployment (percentage of the total labor force) had a higher positive loading into the second principal component. This component is closely related to both the labor economy and the monetary economy. Finally, the third principal component was highly correlated with only one of the original variables that is Inflation. This component resembles the

trade and openness of the economy. Graphically, this can be presented using a biplot and factor analysis and a graph of variables.

G. Hierarchical Regression Model

The hierarchical regression model was conducted to examine the effect of the predictor variables on the predicted variable. Three models were fitted, and the results are summarized in Table VI.

TABLE VI: ESTIMATES OF PARAMETER IN MODELS

| | Model | Coefficients | Std. Error | T value | Sig. |
|---|---|---|---|---|---|
| 1 | (Constant) | 5.086 | 0.373 | 13.650 | 0.000 |
| | PC1 | 1.039 | 0.021 | 48.844 | 0.000 |
| 2 | (Constant) | 5.281 | 0.401 | 13.168 | 0.000 |
| | PC1 | 1.033 | 0.022 | 47.752 | 0.000 |
| | PC2 | -0.041 | 0.032 | -1.267 | 0.212 |
| 3 | (Constant) | 5.549 | 0.388 | 14.286 | 0.000 |
| | PC1 | 1.049 | 0.021 | 49.639 | 0.000 |
| | PC2 | -0.035 | 0.030 | -1.135 | 0.262 |
| | PC3 | -0.302 | 0.111 | -2.722 | 0.009 |

In order to determine the association between the first component (with 15 original variables), the second component with two variables, the third component with only one original variable, and the GDP (dependent variable), the researcher conducted a hierarchical multiple regression analysis as presented in Table VI. As per the R generated output, the equation ($Y = f(PC1, PC2, PC3) + e$) successively becomes;

Mode 1 $$Y = 5.086 + 1.039PC_1 \quad (18)$$

Model 2 $$Y = 5.281 + 1.033PC_1 - 0.041PC_2 \quad (19)$$

Model 3 $$Y = 5.549 + 1.049PC_1 - 0.035PC_2 - 0.302PC_3 \quad (20)$$

In (19), the coefficient of the second component was negative and insignificant. Similarly, in (20), the coefficients of the second component were negative and insignificant, while the third component was negative and significant. After the insignificant coefficients were dropped, the retained coefficients formed (21), (22), and (23).

Model 1 $$Y = 5.086 + 1.039PC_1 \quad (21)$$

Model 2 $$Y = 5.281 + 1.033PC_1 \quad (22)$$

Model 3 $$Y = 5.549 + 1.049PC_2 - 0.302PC_3 \quad (23)$$

The constant terms, which are also the intercepts of the three models, were 5.086, 5.281, and 5.549, respectively. They showed an increasing trend as more components were added to the respective models, as shown in Table VIII. Since the t - values obtained had p - values of 0.000, which were less than the critical value of 0.05, all three constants were significant. The values 5.086, 5.281, and 5.549 are the values of GDP that do not depend on the macroeconomics variables (independent variables). The first component that contained 15 original variables and the gross domestic product was positively and significantly associated because the t value (48.844) attained had a p-value less than the 0.05 critical value (coefficient of X1 = 1.039, p = 0.000). This means that a unit increase in the first principal component holding other components constant would lead to an increase in GDP by 1.039 units, and the influence was significant. Further, after the addition of the second principal component (with two original variables), the results indicated that the first principal component and GDP were positively and statistically significantly related, whereas the second component and economic growth had a negative but the relationship was not statistically significant. This was supported by the t values of 47.752 and -1.267 with their corresponding p values of 0.00 and 0.212, respectively. The coefficient of the first principal component was 1.033, while the second one had a coefficient of -0.041. This implied that a unit increase in the first principal component results in an increase in Kenya's GDP by 1.033 units but GDP decreases by 0.041 units as the second component increase by one unit.

Finally, the third model included all the three extracted principal components as per the original variables. The first component was found to have a significant positive relationship with the GDP while the second one had a negative, but its relationship with economic growth was not statistically significant since its p-value (0.262) was greater than the five percent level of significance. The third component was found to have a significant negative association with the GDP. Additionally, the first component had the greatest influence since its unit increase resulted in a 1.049 unit rise in Kenya's GDP. The second principal component had a coefficient of -0.035 with a p-value of 0.262, indicating that GDP decreases by 0.035 units as per unit increase in the second component, but the influence was insignificant. The coefficient of the last component was – 0.302 with its corresponding p-value of 0.009 and thus having an inverse relationship with the GDP. This shows that an increase in the third component by one unit would lead to GDP significantly reduced by 0.054 units.

TABLE VII: GOODNESS OF FIT OF MODEL

| Model | R | $R^2$ | Adj $R^2$ | SE | F Change | Sig. F Change |
|---|---|---|---|---|---|---|
| 1 | 0.990[a] | 0.980 | 0.979 | 0.147 | 2385.689 | 0.000 |
| 2 | 0.991[b] | 0.982 | 0.981 | 0.146 | 1.605 | 0.212 |
| 3 | 0.992[c] | 0.984 | 0.983 | 0.137 | 7.408 | 0.009 |

The first model in Table VII, has only one independent variable that is the first principal component, had an R-square of 0.98. This indicates that 98% of the variations in Economic growth can be explained by principal component 1, that is, by all 15 original variables contained in PC 1. The p-value was found to be 0.0001, which is less than 0.05 hence implying that the model is significant. This implies that principal component 1 reliably predicts economic growth. After the addition of one variable that is principal component 2, Model two was fitted. The model 2 adjusted R-square was found to be 0.981, implying that 98.1% of variations in economic growth are explained by principal components 1 and 2. Comparing this with model 1, the adjusted R-square has improved, implying that the added variable enhances the model. However, the model 2 P-value was found to be 0.212, which is more than 0.05, implying that the model is insignificant; hence the predictor variables cannot be relied upon in predicting growth. When the third principal component was added, model 3 was fitted, capturing all the 3 PCs. Model 3 adjusted R-square was found to be 0.983, implying that 98.3% of the variation in economic growth is explained by the three PCs. Compared with models 1 and 2, the adjusted R-square for model 3 improved significantly, implying that the addition of component 3 greatly enhances the model. Model 3 P-value was found to be 0.009, which is less than 0.05; hence the model is significant and can be relied upon in predicting economic growth.

TABLE VIII: ANOVA

| | Model | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 51.716 | 1 | 51.716 | 2385.689 | 0.000 |
| | Residual | 1.019 | 47 | 0.022 | | |
| | Total | 52.734 | 48 | | | |
| 2 | Regression | 51.750 | 2 | 25.875 | 1208.991 | 0.000 |
| | Residual | 0.984 | 46 | 0.021 | | |
| | Total | 52.734 | 48 | | | |
| 3 | Regression | 51.889 | 3 | 17.296 | 920.737 | 0.000 |
| | Residual | 0.845 | 45 | 0.019 | | |
| | Total | 52.734 | 48 | | | |

The results show that model 1 had an F statistic of 2385.689 and a p-value of 0.001, which was less than the critical value of 0.05, implying that the model is significant. Similarly, the second model had an F- statistic of 1208.991 and a P-value of 0.001, which is less than 0.05, implying that the model is also significant. The third model had an F-statistic of 920.737 and a P-value of 0.001, which is less than 0.05 hence also being a significant model. A p-value of less than the critical value at 0.05 significance level shows that there is a significant difference in the means of

the variables. Generally, the ANOVA results depict that the differences between some of the means are statistically significant since, for all the models, the p-values are less than the critical value.

There was a decreasing trend of the Mean Square Error (MSE) since Model 1 had an MSE of 51.716, model MSE was 25.875, and that of model 3 was 17.296. Lower MSE values suggest a greater fit. If the main objective of the model is prediction, "MSE is a good indicator of how accurately the model predicts the response, and it is the most relevant criterion for fit". Among the three models, as shown in Table VIII, model one had the highest mean square error while model three had the least, suggesting that it was the best predictor model among the three fitted models.

H. The Best Predictors of Economic Growth

The first component, which consisted of 15 of the original variables, explained up to 73.605% of the total variations, as shown in Table IV. The high variation implies that there was a strong strength of association between the 15 original variables in component one. It also depicts that the components can be used to make better predictions compared to the other two components that had a lower variance explained. As depicted in Table VII, Model 1 had the lowest P-value of 0.001compared two models 2 and 3; hence the independent variable used in model 1 that is component 1, can reliably predict economic growth. Utilizing the Total Variance explained and the P-value, all agree that the first principal component is the best predictor of economic growth. Hence, 15 original variables contained in component 1 are the best predictors of economic growth. These 15 variables depict that the monetary factors of the economy, the trade and openness of the economy with government activities, the consumption factor of the economy, and the investment factor of the economy all have the greatest impact on economic growth.

## V. CONCLUSION AND RECOMMENDATION

A. Conclusion

After the utilization of the PCA technique, three components were extracted and retained. The extracted components explained a total variation of 89.856% in the original data set. This was a high percentage of variation hence depicting the effectiveness of utilization of PCA technique in this study. Since PC 1 had the highest Variance explained (73.605%), it was concluded that the 15 original variables correlated in principal component 1 highly impact economic growth. The 15 microeconomic variables were broad money, households and NPISHs Final consumption expenditure, domestic credit to the private sector, gross capital formation, population total, exports of goods and services, general government final consumption expenditure, foreign direct investment-net inflows, personal remittances received, gross domestic savings, gross national expenditure, aquaculture production (metric tons), imports of goods and services, imports of goods and services, cereal production (metric tons), and official exchange rate. Therefore, the macroeconomic variables associated with the monetary economy, the trade and openness of the

economy with government activities, the consumption factor of the economy, and the investment factor of the economy greatly explain the trend of economic growth in Kenya.

B. Recommendation

It was recommended that when more than 15 variables are being studied, the principal component analysis technique should be utilized to help reduce the dimensionality of the variables and group them into principal components and the hierarchical regression model offers better model building techniques since it captures the partial variance change among the independent variables through the calculation of R-square change from one model to the other.

ACKNOWLEDGMENT

I would like to sincerely thank the Almighty God for the care, knowledge, strength, hope, and patience He granted me during this study. My special appreciation goes to co-authors: Dr. Dennis K. Muriithi, Ph.D., and Dr. Gladys G. Njoroge, Ph.D., for their continuous availability for their technical guidance, valuable and constructive advice during the planning, and development of this research work.

CONFLICT OF INTEREST

The authors declare no competing interests.

## References

[1] Pearson K. Principal components analysis. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1901; 6(2): 559.
[2] Zou H, Hastie T, Tibshirani R. Sparse Principal Component Analysis. Journal of Computational and Graphical Statistics. 2006; 15(2): 265-286.
[3] Corner S. Choosing the right type of rotation in PCA and EFA. JALT Testing & Evaluation SIG newsletter. 2009;13(5): 38-45.
[4] Esmaeili A, Shokoohi Z. Assessing the effect of oil price on world food prices: Application of principal component analysis.
Energy Policy. 2011; 39(2): 1022-1025.
[5] Njoroge E, Njoroge G, Muriithi D. Evaluating Secondary School Examination Results: Application of Principal Component Analysis. Journal of Statistical and Econometric Methods. 2014; 3(2): 31-46.
[6] Boligon A, Vicente I, Vaz R, Campos G, Souza F, Carvalheiro R et al. Principal component analysis of breeding values for growth and reproductive traits and genetic association with adult size in beef cattle1. Journal of Animal Science. 2016; 94(12): 5014-5022.
[7] Field, A. P. Discovering statistics using SPSS (2nd edition). London: SAGE Publication, 2005.
[8] Granato D, Santos J, Escher G, Ferreira B, Maggio R. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for the multivariate association between

bioactive compounds and functional properties in foods: A critical perspective. Trends in Food Science & Technology. 2018; 72: 83-90.
[9] Richardson D, Hamra G, MacLehose R, Cole S, Chu H. Hierarchical Regression for Analyses of Multiple Outcomes. American Journal of Epidemiology. 2015; 182(5): 459-467.
[10] Hussain A, Sabir H, Kashif M. Impact of macroeconomic variables on GDP: Evidence from Pakistan. European Journal of Business and Innovation Research. 2016; 4(3): 38-52.
[11] De Roos A, Poole C, Teschke K, Olshan A. An application of hierarchical regression in the investigation of multiple paternal occupational exposures and neuroblastoma in offspring. American Journal of Industrial Medicine. 2001; 39(5): 477-486.
[12] Brown J. Choosing the right number of components or factors in PCA and EFA. JALT Testing & Evaluation SIG Newsletter. 2009; 13(3): 20-25.
[13] Lever J, Krzywinski M, Altman N. Principal component analysis. Nature Methods. 2017; 14(7): 641-642.
[14] Liu Q, Cook N, Bergström A, Hsieh C. A two-stage hierarchical regression model for meta-analysis of epidemiologic nonlinear dose-response data. Computational Statistics & Data Analysis. 2009; 53(12): 4157-4167.
[15] Field A. An Adventure in Statistics: The Reality Enigma Ed. 1. London: SAGE Publications, 2016.
[16] Marczyk G, DeMatteo D, Festinger D. Essentials of research design and methodology. John Wiley & Sons Inc., 2005.
[17] Fan J, Liao Y, Wang W. Projected principal component analysis in factor models. The Annals of Statistics. 2016; 44(1): 219.
[18] Trustrum K, Fox J. Regression Diagnostics: An Introduction. The Statistician. 1993; 42(2): 201.
[19] Gelman A, Goodrich B, Gabry J, Vehtari A. R-squared for Bayesian Regression Models. The American Statistician. 2019; 73(3): 307-309.