Singular Spectrum Analysis: An Application to Kenya's Industrial Inputs Price Index

B VENKATA RAMANA, D MADHUSUDANA REDDY, R SIVA GOPAL
ASSISTANT PROFESSOR [1] , PROFESSOR[2,3]
bvramana.bv@gmail.com, madhuskd@gmail.com, sivagopal222@gmail.com
Department of Mathematics, Sri Venkateswara Institute of Technology,
N.H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

Abstract

The time-related features and future trajectory of any given time series may be better understood with the help of time series modelling and forecasting methods. The use of classical models without data transformation is hindered since the majority of financial and economic time series data does not adhere to the stringent assumptions of data normalcy, linearity, and stationarity. Because Singular Spectrum Analysis (SSA) is data-adaptive and non-parametric, it does not need to adhere to the same stringent assumptions as classical approaches. Using data collected from January 1992 through April 2022, this study examined the relationship between SSA and Kenya's monthly industrial inputs price index using a longitudinal research approach. In an effort to promote competitiveness and growth in Kenya's manufacturing sector, one of the country's manufacturing priorities since 2018 has been to lower the prices of industrial inputs. By 2022, the anticipated Manufacturing Value Added for Kenya was supposed to reach a whopping 22 percent. The research found that the SSA yields better accurate predictions (L = 12, r = 7, MAPE = 0.707%). According to the 24-period estimates, the cost of industrial inputs is expected to be reduced in the post-industrial agenda, although it is still high compared to 2017. So, we need to bring down the costs of industrial inputs to a reasonable level.

Keywords: Singular value decomposition, industrial input price index, and singular spectrum analysis.

I. Introduction

When a country's economy shifts from being dependent on agriculture to producing things, it is known as industrialization. Industrialization guarantees maximum utilisation of resources, increases international commerce, and generates job possibilities, all of which contribute to long-term economic growth and development.within the domain of economics.The manufacturing sector has the biggest employment multiplier impact compared to other sectors, according to empirical studies[1, 2].Consequently, creating

The transition to skill-intensive manufacturing may help economies generate a large number of jobs in non-tradable industries [2]. In 2018, the industrialised nations with the greatest manufacturing value added (MVA) as a proportion of GDP were Germany (21%), France (21%), Japan (21%), and the United States (10%) [3]. In 2018, China had 28% MVA and South Africa (SA) had 12% among the recently industrialised nations, which includes Brazil, Mexico, and China [3]. The service sector accounts for the largest portion of Kenya's GDP, followed by agriculture, forestry, fisheries, manufacturing, and industry (building). The four-sector MVA in 2016 was 44.8% [4], 31.1% [5], 17.9% [6], and 9.3% [7] correspondingly. Each of these

MVAs was 43.2% in 2019, 34.1% in 2018, 16.1% in 2019, and 7.5% in 2018. Compared to the construction (-10.056%) and service (-3.57%) sectors, manufacturing's MVA declines by a larger amount, or around 19.36%, from 2016 to 2019. Although manufacturers find Africa to be the most cost-effective place because to its abundant and inexpensive labour force, many Kenyan firms are struggling with rising input costs, high taxes, and power prices. It is estimated that at least 80% of the raw materials used in Kenya's industrial industry are imported. Industrial competitiveness in Kenya is therefore strongly related to the cost of imported inputs. From 8.4% in 2017 to 15% in 2022, MVA is expected to rise, as per the KAM priority agenda [8]. An objective of the first pillar of improving manufacturing, which is levelling the playing field and fostering competition, is to reduce the cost of imported industrial inputs [8]. When formulating or evaluating policies, it is usual practice to examine the development of a series using time series analysis methods. Parametric and non-parametric analytical models are the two main categories. The characteristics of the data dictate which of the two main categories of models to use. The data must be normal or have a homogeneous variance in order to use parametric approaches. Depending on the statistical test used, two possible courses of action might be considered in the event that the facts contradict the assumption. The non-parametric options may be used, or the data can be transformed to fit the previous restrictive assumptions using linear or non-linear transformations, or outliers can be removed as needed. When looking at time series, the distribution and stationarity of the data are the most obvious concerns. It is possible to make use of data that is normal or stationarity using classical time series models like GARCH and ARIMA. However, these assumptions are not met by the majority of economic and financial time series data. The trend component that time-related elements impart is another distinctive property of time series data. In addition, when it comes to estimating and specifying the ARMA class of models, level shifts become a difficulty [9]. When employing classical time series approaches, data modifications using methods like differencing are necessary to fix violations of the stationarity requirements. Such data modification, say Hassani and Mahmoudvand [10], wipes out any series' temporal information. Therefore, parametric time series models may not be able to get financial and economic time series predictions right.

defined by structural fractures, non-stationarity, and divergence from normalcy brought about by external shocks including changes in technology, government policies, and consumer tastes [11]. Other factors, particularly in the financial markets, include announcements, news, and insider knowledge, all of which may lead to unexpected and fast changes in stock prices.

When non-parametric models like singular spectral analysis (SSA) are used, the parametric assumptions about normality, stationarity, and linearity are not relevant. Applying PCA to time series data allows for the creation of the SSA. In order to decompose a trajectory matrix into its constituent parts, SSA employs matrix algebra. To simulate the trend, oscillatory, and noise components in a certain time series, the components are grouped and rebuilt according to their periodicities [12]. In contrast to the Fourier method, which relies on sine and cosine functions, or mother wavelets, SSA signal decomposition and reconstruction use entirely data-adaptive basis functions, specifically the eigenmode of the trajectory matrix [13]-[15]. In addition, SSA is superior to other non-parametric methods since it can describe any time series with only two parameters: the window length (L) and the number of components (r).

II. LITERATUREREVIEW

Numerous empirical investigations have shown that SSA is a powerful tool for modelling and prediction. Numerous fields have made good use of SSA, including paleoclimatology [20], economics [19], medicine [17], image processing [18], economics [19], and weather forecasting [16]. This is because SSA offers practical benefits over classical models. According to the majority of empirical investigations, the SSA approach outperforms traditional models in terms of predicting ability.As an example, in a region of Venezuela that relies on wind power, [21]compared the SSA, HWES, and SARIMA models with monthly electric load demand (in MW). The researchers used a cross-validation method, dividing the whole data set into a training set (n1 = 228) and a test set (n2 = 12). The training set was used to fit the model parameters, while the test set was used to evaluate the model's performance. In order to achieve greater component separability, the research picked a window length (L) of 108 using the formula L= (T-s)/2, where N = 240 and s is the periodicity of the series (i.e., s = 12). The research concluded that seven important eigentriples were used throughout the rebuilding phase. The findings demonstrated that SSA (Mean Absolute Percentage Error, or MAPE) = 1.06%) outperformed the Additive HWTES (MAPE = 1.25%) and ARIMA (2,1,0)(2,0,2)! (MAPE = 1.15%) models in predicting monthly electric load demand, according to the reduction of in-sample prediction errors. In creating 12-step forward projections, the SSA model (MAPE = 1.26%), the Additive HWES model (MAPE = 2.28%), and the *ARJMA* (2,1,0)(2,0,2)! model (MAPE = 1.86%) all performed similarly throughout the in-sample evaluation period. This consistency was maintained during the out-sample assessment as well.

In reference [22], the SSA technique, ARIMA, ES, and NN models were tested for their susceptibility to the Great Recession. Authors segmented export and import data into three time periods using US recession dates, data visualisation, and the structural breaking points test proposed by Bai and Perron (2003). There were three distinct time periods covered by the import statistics: before (January 1989–July 2008), during (August 2008–April 2009), and after (May 2009–December 2011) the global economic slump. From January 1989 through June 2008, throughout and after the recession (July 2008–January 2009), and from February 2009–December 2011, the statistics included exports. Using cross-validation, the authors used a training set that consisted of two-thirds of the data collected before and after the recession. A third was used to assess the precision of the predictions made for1,3,6, and 12 months in the future. The authors determined that SSA is less affected by U.S. trade recessions compared to optimised ARIMA, ES, and NN models by minimising the Root Mean Square Error (RMSE), the direction of change (DC) criterion, the modified Diebold-Mariano test, and the t-test. After controlling for other factors, the recession revealed that SSA estimates for US imports and exports were the most accurate.

No news. Take U.S. imports as an example. The SSA prediction outperformed the ARIMA, ES, and NN models before the 2008 recession by an average of 11%, 9%, and 51%, over all step forward horizons, respectively, according to the average Relative RMSE (RRMSE). The SSA's recession period projection for 2008 was 40% better than ARIMA's and 6% better than HW's model. With 34% better predictions than the ARIA model, 42% better than the ES model, and 27% better than the NN, SSA proved its robustness in the analysis period after the recession. It seems to reason that the ES and NN couldn't fit a good model since their sample sizes were too tiny for each time period. The authors suggest using the approach for more accurate long-term projections and state that SSA is an effective noise filtering strategy.

Using monthly U.S. tourist arrivals data from 1996 to 2012, the accuracy of SSA, ARIMA, ES, and feed-forward Neural Network (NN) forecasts was evaluated in the study by [23].Using cross-validation, the authors trained the model using 355 observations from the sample (n1 = 355) and tested the accuracy of the forecasts over a specified time period using 356 observations (n2 = 356). Prediction errors caused by varying forecasting horizons were minimised by using varied values of L and r in the model. In every h-(1, 3, 6, 12, 24, and 36) step-ahead prediction, SSA beat the other three models. As an example, SSA had a better performance than SARIMA (2,0,1) (1,1,2) (MAPE = 14%), ES (α = 0.86, σ = 0.1231) with multiplicative seasonality (MAPE = 17%), and NNAR (2,1,1) (MAPE = 23%). In order to predict the daily demand for ambulance services supplied by the Welsh Ambulance Service Trust (WAST) in Wales, researchers assessed the efficacy of three different methods: SSA, ARIMA, and HWES.The information spanned the months of April

the first day of 2005 until the last day of 2009. With SSA (L = 581, R = 14), the ambulance demand series was best fitted to a training set consisting of the first 51 months of data (N = 1,552). In terms of minimising the root-mean-squared error (RMSE), SSA performed better than HWES method for both in-sample forecasts (RMSE SSA = 6.19, RMSEHWES = 6.37), and for forecasts made h-steps ahead of time (7-day: SSA=42.20, HWES=45.46; 14-day: SSA=42.86, HWES=47.47; 21-day: SSA =43.87, HWES =48.32; 27-day: SSA =45.46, HWES = 51.14). Based on the study's results, SSA is a good choice for predicting ambulance demand in Wales since it remains stable across extended projection horizons.
When compared to conventional models like the ARIMA class of models, the SSA technique shows superior forecasting abilities across a variety of sectors, economies, time periods, and frequencies. The present research contributes to the current body of knowledge on SSA by applying it to the industrial inputs price index in Kenya.

III. METHODOLOGY

A. Study Plan A longitudinal research approach was used in this study. Longitudinal analysis is a subset of time series analysis that examines a variable's spatial and temporal changes via the examination of consistent historical data collected several times [25]. Examining the time-related characteristics of Kenya's industrial inputs price index is the focus of this research.Seasonal, trend, cyclical, and error components may all be shown using time series. Table B. Since this is a study of univariate time series data, no sampling method was used. Based on data extracted from the IMF website, the research analysed a 26-year sample of Kenya's monthly industrial inputs price index data, covering the period from January 1992 to April 2021. Given that the SSA is applicable to even lower sample sizes than 120 [12], 352 data points are deemed suitable for the time. This index is a combination of the agricultural raw materials index and the basic metals index. C. Analysing Data The statistical software R was used for data analysis. The SSA modelling process consisted of many steps, including series embedding, series decomposition using SVD, component grouping, time series additive component reconstruction, and series forecasting [12]. Below, we will go over the steps.

Introduction: Embedding $Y\# = [x\$, x!, x", \ldots, X\%]$ is the series to be defined. The process of embedding involves using the past values of $a\#$ to derive a vectory matrix (Y) with dimensions $L \times (N - \mathbf{O} + 1)$ from the time series $\{a\#\}$, where N is the length of the series and L is the length of the window. When determining the separability of various eigenmodes, the most important parameter is the window length.A window's length cannot exceed $2 \le \boldsymbol{X} \le N/2$.

After choosing the best L, the series $a\#$ is transformed into a $\mathbf{O} \times K$ trajectory matrix Y, as shown in (1).

$$Y_{L\times K} = T(X_t) = [Y_1, Y_2, Y_3, \dots, Y_K]^T \quad (1)$$

Where;

$$Y_i = [x_{i-1}, x_i, x_{i+1}, \dots, X_{i+L}]^T \quad (2)$$

Step2:DecompositionStage

The SVD is a crucial process of SSA, which helps identify the periodicities of different Eigen/principal modes ranked by their significance or share of variance explained. The SVD theorem states:

$$A_{t\times y} = U_{t\times t} S_{t\times t} V_{t\times y} \quad (3)$$

Where columns of the matrix $U_{txt}$are the left singular vectors (row–component correlation); S is a diagonal matrix of singular values, the rows of matrix V are the right singular vectors (column- component correlation) [26]. The $U_{t\times t}$consists of the eigenvectors of AA^T whereas the eigenvectors of A^TA yields the columnsof$V_{t\times y}$.Theeigenvaluesandeigenvectors of $AA^T$and A^TAareobtainedusingtheeigen decomposition [27]. In the current study, the resultant trajectory matrix Y is decomposed using SVD into three parts as expressed in (4).

$$Y = \sum_{i=1}^{d} \sqrt{\lambda_i}\, U_i V^T{}_i \quad (4)$$

Theabovenotation can beexpressedas additivecomponents in (5):

$$Y = Y_1 + Y_2 + \cdots + Y_d \quad (5)$$

Where,λ+representsthesingularvaluesofmatrix Yarrangedindecreasingorderofmagnitude corresponding to the i#2 eigenvector, U+ arethe left singular vectors, V+ arethe right singular vectors. In univariate series, of interest is to get eigentriples (Bλ+U+V+) of the SVD. Step 3: Grouping

Each eigentriple is a vector, which, before grouping, can be seen as a possible component. The periodicity and contribution of each Eigen component (eigenvector) depend on the size of their respective eigenvalues. Major/principal components carry a larger weight, whereas more unwanted signals (noise) have low variance and come last [12]. Hence, it is possible to identify the Eigen components corresponding to noise. The contribution of the eigenvector can be expressed in terms of the proportion of variance it explains. The share of each eigenvector ϕ+ is obtained using (6).

$$\phi_i = \frac{\sqrt{\lambda_i}}{\sum_{i=1} \sqrt{\lambda_i}}$$

Where $\sqrt{\lambda}+$ is the square root of the eigenvalue $E+$ associated with eigenvector $V+$.
Understandingtheproportionofvarianceisvitalbecausemeaningfulvectorsthatexplainasubstantialvariationintheoriginalseriesareusedtoreconstructthetimeseriescomponents, whereasinsignificantvectorsaretreatedasnoise.Thestudyretainedalltheeigenvectorsthatyieldacumulativevarianceexplainedof100% toensurenovitalinformationfromthedatais lost.
The components are grouped as pairs if they have the same periodicity [12]. Insignificant Eigen triplets that correspond to low eigenvalues as they resemble the unwanted noise are dropped. If, for example, there exist two such groups given by $G!$ ={ $Y!$, $Y"$,

$$Y_{rec} = G_1 + G_2 + \cdots + G_r = \sum_{i=1}^{r} \sqrt{\lambda_i}\, U_i V^T{}_i \qquad (7)$$

$Y4$}, $G2$ ={Y), $Y5$} among d triplets, then; $Y678$ = $G!$ + $G"$ + $G6$ where $G6$ ={ $Y9$, Y:,

$$X_{N+h} = x_1, \dots, x_{N+h} \qquad (9)$$

… , $Y0$} is the residual signal constituting the noise. The grouping $G!$ + $G"$ + ⋯$G6$ yields a reconstructed trajectory matrix ($T6$), as expressed in (7).

Theseparabilityoftheprincipalcomponentscanbedeterminedusingthe Weightedcorrelation ($W8$) computed using (7) [28].

$$W_c = \frac{(Pc^1,}{||Pc^1||_w}$$

Where $||Pc^1||_w = \sqrt{(Pc^1, Pc^1)_w}$:

$$(Pc^1, Pc^1)_w = \Sigma_l^l$$

The weighted correlation quantifies the degree of separability among two reconstructed components. Let X= and X> be two reconstructed time series. The two series are said to be separable if w-correlation (X=, X>) is zero. Contrarily, two or more components are grouped if the resultant w-correlation (X=, X>) is relatively high or significant.
Step4:ReconstructionStage
The next step is the reconstruction of the deterministic component of the time series of length N by applying diagonal averaging on the Trajectory matrix $T6$. The SVD process is reversed through a newly reconstructed trajectory matrix to a deterministic univariate series. Diagonal Averaging provides an optimal way to return to a single time-series data.
Step 5: Forecasting
The forecasts are obtained from the reconstructed series. The h-step ahead predictor model in SSA is defined using (8) [20].

Forecastingshouldsatisfyrecurrentlinearformulaein(10):

$$x_{i+d} = \sum_{k=1}^{d} \alpha_k x_{i+d-k}\,, 1 \le i \le N-d \qquad (10)$$

Where d is the dimension of the series and $\alpha<$ are the coefficients. Notably, d is the rank of Y, which is equal to the rank of $YY*$.

A. ModelsEvaluationCriterion

The current study employed the Mean Absolute Percentage Error (MAPE) to compare the performance of SSA and the two HWTES models, namely, the additive and multiplicative. The formula in (11) gives the MAPE.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - F_t}{X_t}\right| \times 100 \qquad (11)$$

The provided forecasting model is used to fit the actual data, which is represented by $X\#$.

B. Examinations Conducted After the Diagnosis!

Analysing the residual distribution was one of the post-diagnostic testing. A good model basically has residuals that are white noise (random), distributed properly, and not autocorrelated. Its unpredictability may be tested using a visual plot. It is anticipated that the model residuals will exhibit a patternless, completely random process in this case. A normal distribution should be checked in the present study's examination of the models' in-sample residuals using a histogram, which brings us to our second check. Lastly, we assume that there is no residual autocorrelation in our test. The ACF plot was used to test the hypotheses that there is no autoregressive dependency among the residuals.

.

IV. RESULTSAND DISCUSSION

V.Section A. Initial Evaluation

A thorough comprehension of the data's distribution and development is crucial for selecting a time series model. Data visualisation, descriptive statistics, and series decomposition were therefore used as preliminary analyses in the present research. From January 1992 to April 2021, the time series graphic in Figure 1 shows the development of the monthly industrial inputs price index data for Kenya. There was relative stability in the price index during the '90s until 2002/03. From then, it rose steadily until 2007.The severe decline that occurred during the 2008–2009 economic recession, which was caused by the financial crisis, is clearly visible. A drop occurred from 2011 to 2016 because the government changed hands smoothly after the 2012 elections. Despite Kenya's industrial plan aiming to reduce the cost of industrial inputs in 2017, the index of industrial input prices continued to increase [8].Because of this tendency, we may understand why Kenya's MVA fell from 9.3% in 2016 to 7.5% this year [7]. The series' periodic variation demonstrates that the industrial inputs price index is sensitive to political and economic shocks on a worldwide scale, such as the 2008–2009 financial crisis and the 2019 Coronavirus Disease (COVID-19).A statistically insignificant

External shocks may cause a time series to become noisier or even non-stationary by inducing the seasonality component. With ADF= -2.1324, Lag order= 7, and p-value= 0.5209, the Augmented Dickey-Fuller (ADF) test confirmed that the data is non-stationary and suffers from a unit root issue.

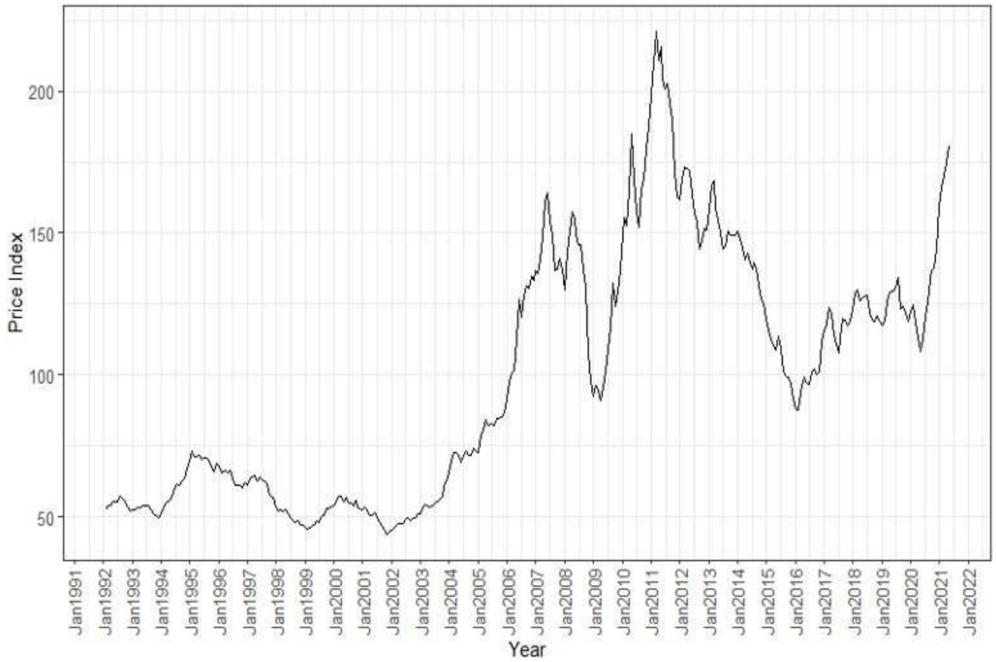

Fig.1:Monthlyindustrialinputspriceindexdata(1992M1–2021M4).Note:Thebaseyearfor the 2016 is the year's series

During the research period, the average industrial inputs price index was 99.29 (SD = 44.09), as shown in Table I. The data is almost symmetrical to the mean, since the metrics of dispersion are close to zero (Skewness = 0.490, Kurtosis = -0.085). The data from the industrial inputs index does not follow a normal distribution, according to the Shapiro-Wilk (S-W) test (W = 0.9116, p = 0.000). When the Kurtosis test returns negative, it means that the series' distribution is platykurtic, which means that the data is flatter than normally distributed data, has a wider peak, and tends to spread about the mean.

TABLEI:DESCRIPTIVESTATISTICS

	N	Min	Mean	SD	Max	Skewness	Kurtosis
IIPI	352	43.78	99.29	44.09	221.18	0.490	-0.085

Note. IIPI: Industrial Inputs Price Index.

The chosen non-parametric models are acceptable since the data is non-stationarity and non-normal. Seasonality and trend patterns are often shown by the series. Because they separate the trend and seasonality components from the irregular variation, smoothing methods like SSA may improve findings [29].

Section A: The SSA Approach

Using the series of industrial inputs price indices, the following is an empirical discussion of the phases of SSA:

Introduction: Embedding

The first step in using structural vector descent (SVD) on a one-dimensional series, such the industrial inputs price index series, is embedding, which involves transforming the univariate series into a matrix. The so-called Hankelization method entails combining the lagging values into a trajectories matrix with dimensions $\mathbf{O} \times K$. At this point, finding the ideal window length (L) is the most important job. The eigenmodes that are produced are separable to varying degrees depending on the magnitude of L. There is no universally ideal window length since distinct series may display various time series components at different points in time. Furthermore, time series frequency differs between studies. Finding the ideal window length is therefore beneficial to the researcher. The interval $2 \leq a \leq N/2$ is proposed by Golyandina et al. [12]. The use of the Hadamard transform in time series reconstruction outperforms that of ordinary SSA, as recently discovered in [30], which led the authors to suggest ways to enhance time series analysis, especially when utilised Lin the embedding step of SSA. Accordingly, L should be chosen from the maximum-order Hadamard transform ($\mathbf{b}$2) that satisfies 22<N. In this investigation, the points chosen for the sample (352) fall somewhere on the interval (256, 512] or (2@, 2A). This means that $\mathbf{O}\prod$ = 22,! = 2: = 128 is the maximum size of the time series embedding, and therefore h = 8. A series of 12s between 12 and 180 (NB: N/2 = 176) was supplemented with 128 as potential window lengths since, according to [31], window lengths that are divided by the period provide superior separability. In order to look ahead, we calculated the prediction errors for each window length (L) while taking the ideal number of significant eigenvalues (r) into account (Fig. 2).

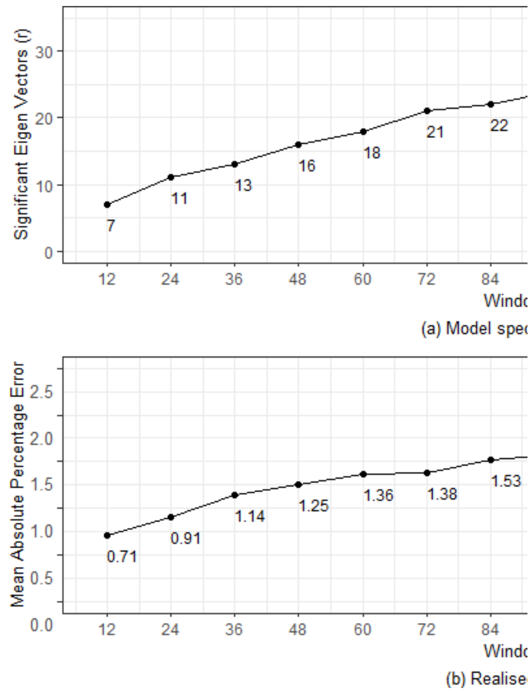

Fig.2:Wind owlength selection.

The MAPE shows that the prediction errors grow with increasing window length. A windowlength of 12 - the series' frequency - with an ideal r of 7 (MAPE = 0.71%) was determined for Kenya's industrial input price index. Hence, SSA(L=12, r=7) is the appropriate specification for the IIPI SSA model. The findings are in agreement with previous empirical research, such as [11], which discovered that as L is increased, the RRMSE for reconstructing a simulated sinusoidal noisy series gradually increases. The best-fit window length reduction of the prediction errors (MAPE = 9%) was determined to be 15 with associated 6 eigenvectors using a sample of 355 observations of monthly US tourist arrivals data as the training set [23]. Researchers have discovered few meaningful eigentriples to rebuild the series in studies that have used a significantly longer window length. This suggests that a higher dimensional matrix is unnecessary since it represents data redundancy, which might be the source of overfitting. To illustrate, in a Venezuelan area that relies on wind power for electricity generation, [21] used 228 samples of training data on monthly electric load demand (in MW) and identified 7 important eigentriples that were utilised during reconstruction. This value is similar to the standard formula of $L = (N - 12)/2$. The desired tiny value of L, which is 12, is also indicative of the series' character. Since a shorter window length is more suitable for extracting the trend component—which accounts for a greater amount of variance in the studied series of Kenya's industrial inputs price index—we have chosen to use it. This aligns with the expectation of [12].

Step 2: Decomposing into its component parts
The primary goal of support vector descent (SVD) is to partition the trajectory matrix into smaller parts that can store important information from the larger matrix. The chosen window length is used to embed the data into a trajectory matrix, Y340x12, which is further deconstructed using SVD, as shown in (12).

The $(U+)$ and the $(V+)$ are the and left and right singular vectors $(V+)$whereas the singular values (scalars) $(\sqrt{\lambda}+)$ denote the scaling. Of interest is to get the collection B$\lambda+U+V+$ known as the eigentriples, which are donated by $F+$. The SVD decomposition of the embedded matrix results in 12 leading eigentriples (F)5$×!") (Fig. 3) Notably, the eigenvectors are usuallyarranged in descending order based on the magnitude of their singular values.

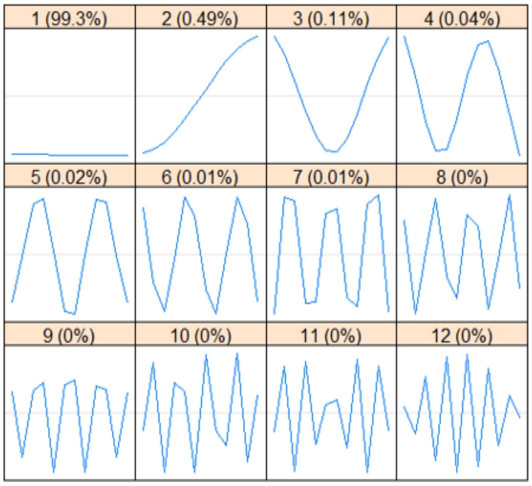

Fig.3:LeadingeigenvectorsfromSVDdecomposition.

It seems that the time series components of the industrial inputs price index series can be sufficiently explained by the first seven eigentriples, as the curve levels out after the seventh eigenvector. The choice is in line with the 100% variance explained by the generated eigenvectors as seen in Figure 5. Therefore, eigentriples 1–7 were used for signal reconstruction, whereas the other eigenvalues are considered noise.

Parts 3 and 4:Reconstruction and grouping

Finding a signal in the time series characteristics of index series of industrial input prices is made easier by the final two phases of the SSA approach. The seven eigentriples may be considered independently, however they must be grouped according to their connection. When arranging Eigen triplets, periodicity of the irrespective eigenvectors is the deciding factor [12].When separating and identifying harmonic components, it is helpful to keep in mind that the singular values of the two eigentriples in the series are quite near to each other. The eigentriples that correspond to the harmonic components of the series may be identified by analysing the pairwise scatterplots of the singular vectors (Fig. 5).

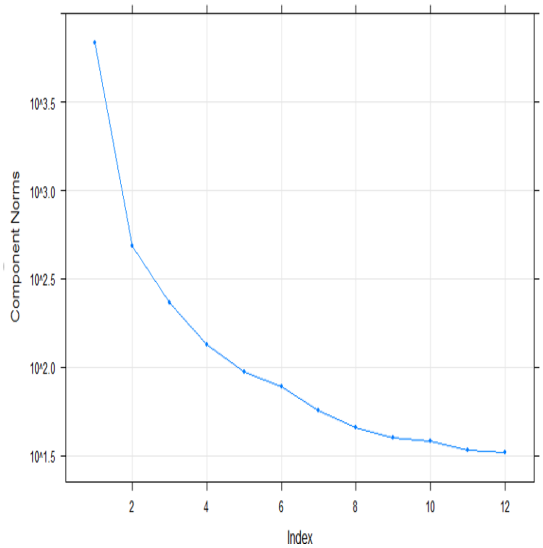

Fig.4:Scree-plotofcomponentnormsagainsttheeigentriples.Note.$Norm$!='Yu'"=
√Yu"#Yu"=√λ!=$singularvaluei$

It seems that the time series components of the industrial inputs price index series can be sufficiently explained by the first seven eigentriples, as the curve levels out after the seventh eigenvector. The choice is in line with the 100% variance explained by the generated eigenvectors as seen in Figure 5. Therefore, eigentriples 1–7 were used for signal reconstruction, whereas the other eigenvalues are considered noise.

Parts 3 and 4:Reconstruction and grouping

Finding a signal in the time series characteristics of index series of industrial input prices is made easier by the final two phases of the SSA approach. The seven eigentriples may be considered independently, however they must be grouped according to their connection. When arranging Eigen triplets, periodicity of the irrespective eigenvectors is the deciding factor [12].When separating and identifying harmonic components, it is helpful to keep in mind that the singular values of the two eigentriples in the series are quite near to each other. The eigentriples that correspond to the harmonic components of the series may be identified by analysing the pairwise scatterplots of the singular vectors (Fig. 5).

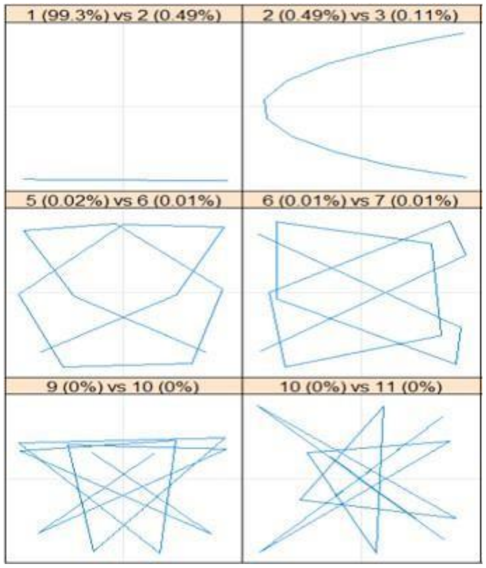

Fig.5:Pairwisecorrelationofeigentriplesfrom SVD.

The geometric figures help visualize the degree of correlation between any two close eigenvectors. The first two vectors have nearly no correlation, whereas F2, F3, and F4 are somewhat correlated. The other geometric shapes denoting the pairwise correlation between the subsequent eigenvectors have no apparent pattern. Grouping using a scatter plot is subject to visual bias. As such, the separability of the twelve possible eigentriples can be quantified using a w-correlation (w-cor.) matrix. The matrix in Fig. 6 depicts weighted correlations between 12 eigenvectors as possible components to reconstruct the time series. Each shaded cell $(F+, FC)$ symbolizes the correlation between eigenvectors i and i, coded from black (w-cor. = 1) to white (w-cor. = 0). Similar components have a strong correlation (darker shading) and are grouped. Conversely, components with a lighter to the white background are linearly independent hence separable. For instance, the first eigenvector (F1) (now component 1) is separable from the rest and makes up the first reconstructed principal component (RC1). In the same logic, F2 to F3 constitute RC2, F4 to F7 make up RC3. The three components correspond to the time series components, namely level (RC1), trend (RC2), and seasonality (RC3). As initially indicated that the first seven eigenvectors explain 100% of the industrial inputs price index variance. Thus, the remaining eigentriples (F8 to F12) form the residual signal or noise.

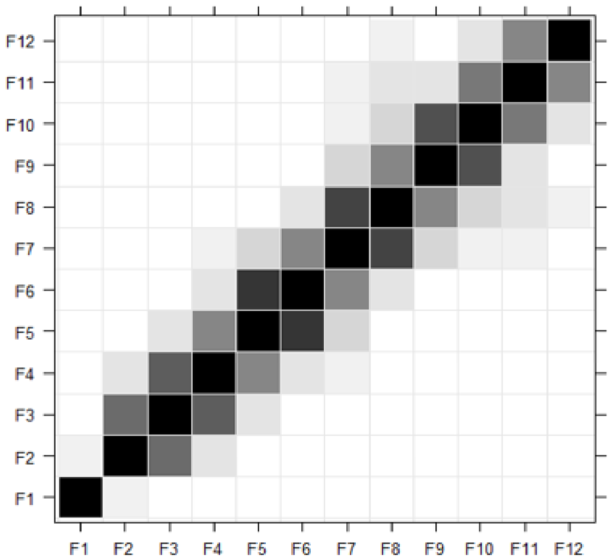

Fig.6:W-correlation matrix.

In Fig. 7, we can see the three rebuilt parts. The total of RC1, RC2, and RC yields the original series since SSA is additive. For time-series data, the RC1 stands for the ubiquitous periodic averages or fundamental nominal changes in the level [9], [32]. Time series exhibit changes that are in keeping with their long-term trends, which might culminate in the formation of the so-called cyclical component of a series. Somehow, the level was low and stable before 2003. During this time, the trend component shows that prices are steady, oscillating about zero. According to [33], the agricultural sector's material and service input costs rose sharply between 2003 and 2006, which contributed to the overall increase in the cost of agricultural goods. The increase in the prices of these goods, which are used in the manufacturing process,

industry, which may be linked to the increasing price of raw materials used in production, such fuel, seeds, and fertiliser. The cost of imported raw materials including rubber, copper, aluminium, and iron ore is heavily influenced by import tariffs, railway development charge, and import declaration fee (IDF) at rates of 1.5% and 2%, respectively [34]. Since the charges are constantly about the same, they don't really change the levelofiberies.All things considered, these factors have been linked to the rising trends and changes in the cost of industrial inputs index. Since the trend's direction and size might alter over time, the RC2 follows suit.After 2005, the trend component of the industrial inputs price index began to fluctuate erratically, demonstrating that the index is dynamic and vulnerable to changes brought about by abrupt shocks like the post-election violence of 2007 and the subsequent precipitous drop in input costs that coincided with the 2007–2008 financial crisis, both of which reduced producers' incentive to in the amounts of consumption being unpredictable. Input costs fall as a result of lower demand for raw commodities.

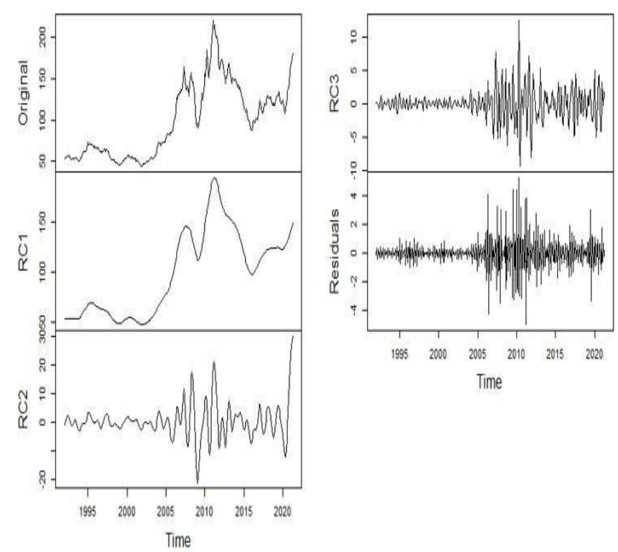

Fig.7:Reconstructed series.

Impact on input costs from the recent COVID-19 epidemic became apparent in the middle of 2020 and continued until 2021. Compared to previous shocks, the pandemic had the largest effect, as seen by the reported strong and protracted growing trend. Raw material supply networks, both local and imported, were thrown into a loop when the COVID-19 pandemic hit almost every country. Closing regional and international borders was one of the measures used during the COVID-19 pandemic to control the supply of raw materials. You may use the seasonal component as a tool to find potential shocks that happen at regular intervals, such quarterly, semiannually, or monthly. When looking at RC3 visually, there is no discernible pattern to the series' seasonality. In the aftermath of unexpected shocks like the 2007 post-election violence, the 2008/2009 financial crisis, and the current COVID-19 outbreak, a seasonal component develops.There has been a general upward tendency in Kenya's industrial inputs price index, however the overall trend is unstable. The findings of the present investigation show that SSA is suitable for extracting the time series components, much as the evaluated studies.

A. Post-test evaluations

Examining the residual distribution was one of the post-diagnostic tests done to ensure the suggested approach was adequate. The histogram and Figure 8 reveal that the residuals are roughly regularly distributed and look like white noise. Finally, the partial autocorrelation coefficients of the residuals up to lag 36 are shown via the Autocorrelation Function (ACF). It is not troublesome if certain autocorrelation coefficients above the 95% significance range.

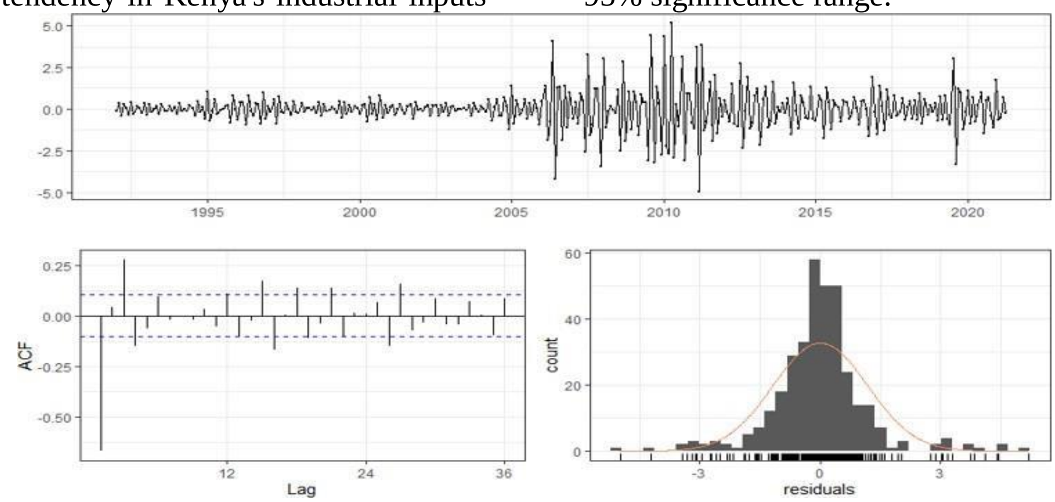

Fig.8:SSAmodel’s residuals. Step5: Forecasting

Journal of Management & Entrepreneurship
ISSN 2229-5348

UGC Care Group I Journal
Vol-9 Issue-01 April 2020

The ultimate goal of the reverse process from diagonal averaging is to obtain a univariate time series of length N. The sum of the four reconstructed components yields the fitted time series or in-sample forecasts (Fig. 9). The in-sample forecasts mimic the original series with minor deviations hence a good fit for the data. One of the most interesting findings is that the 24-step ahead estimates cannot be affected by an extended trend component in the last period of the estimation data. The model's forecasted curve elbows back after the second step-ahead predictions and mimicked the typical behavior of a time series.

Fig.9: SSA(L =12,r=7) forecasts.

The 24-step forward predictions from the fitted SSA (L = 12, r = 7) model are shown in Table II. From June 193.65 in 2021 to June
178.43 in 2022, the numbers show a declining trend of around 5.64%.From 178.79 in July 2022 to 184.71 in April 2023—a 3.52% increase—the abrupt decline is projected to be followed by an upward trend.

TABLEII:SSAFORECASTS

Year	Jan	Feb	Mar	Apr May	Jun	Jul Aug	Sep	Oct	Nov	Dec
2021				189.1	193.7	11 99 20. .3 5	190.6	192.4	191.9	188.2
2022	184.6	183.9	184.7	11 88 30. .8 8	178.4	11 78 80. .6 8	181.2	180.3		181.6
2023	183.9	18.9	146	1						

VI. CONCLUSION

Economic officials cannot do their jobs well without time series analysis and forecasting. Models that provide a better fit for the data and correctly anticipate the future path of a particular series are necessary for forecasting. Since most financial and economic time series do not adhere to the non-stationarity assumptions made by parametric approaches, non-parametric models like the SSA methodology are often better. In addition, while SSA is both a model for predicting and a method for filtering data, it is unlike conventional models in many respects. In light of external and internal shocks like the 2008–2009 financial crisis, recurrent political instability around election time, and the recent COVID-19, the present study showed that SSA methodology clearly differentiates the time-series properties of Kenya's industrial inputs price index (level, trend, and seasonality). Industrial input costs have risen sharply, with the shocks playing a role. Accurate predictions were given by the fitted model. The industrial inputs price index is expected to recover in the final quarter of 2021, according to the 24-period projections using the SSA(L=12,r=7) technique. Then, starting from the third quarter of 2022, there will be another upward trend. Even with the post-industrial agenda's focus on lowering the cost of industrial inputs, the index is higher now than it was in 2017. As a whole, the cost of production rises due to the high cost of industrial inputs, which discourages investment in the industrial sector and slows its expansion. Second, as industry investors struggle to hold on to their profit margins, cost-push inflation is a real possibility. So, it's important to work together to bring down and stabilise the costs of industrial inputs to a sustainable level so that Kenyan manufacturers can be more cost-competitive.

References

[1] BivensJ.(2003).UpdatedemploymentMultipliersfortheUSEconomy.Report,Economic Policy Institute, August.

[2] BergerT,ChenC,FreyCB.Cities,IndustrializationandJobCreation:Evidencefrom Emerging Countries. [Internet] mimeo, Oxford Martin School. 2017. Available from: https://www. oxfordmartin. ox. ac. uk/downloads/JEG_Manuscript. pdf.

[3] WorldBank.Manufacturing, valueadded (%of GDP)|Data[Internet]. Data.worldbank.org. 2021[cited19September 2020]. Availablefrom: https://data.worldbank.org/indicator/NV.IND.MANF.ZS

[4] WorldBank.Services,valueadded(%of GDP) - Kenya |Data [Internet]. Data.worldbank.org.2021[cited14September2020]. Availablefrom: https://data.worldbank.org/indicator/NV.SRV.TOTL.ZS?locations=KE

[5] WorldBank.Agriculture,forestry,andfishing,valueadded(%ofGDP) -Kenya|Data [Internet]. Data.worldbank.org. 2021 [cited 14 September 2020]. Available from: https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=K

[6] WorldBank.Industry(includingconstruction),valueadded(%ofGDP)-Kenya|Data [Internet]. Data.worldbank.org. 2021 [cited 14 September 2020]. Available from: https://data.worldbank.org/indicator/NV.IND.TOTL.ZS?locations=KE

[7] WorldBank.Manufacturing, valueadded(%ofGDP) -Kenya |Data[Internet]. Data.worldbank.org. 2021 [cited 14 September 2020]. Available from: https://data.worldbank.org/indicator/NV.IND.MANF.ZS?locations=KE

[8] Kenya Association of Manufacturers. 2018-Manufacturing-Priority-Agenda: Sparking Kenya’sIndustrialTransformationforJobCreation[Internet].2018[cited14September 2020]. Available from: https://kam.co.ke/kam/wp-content/uploads/2018/02/2018-Manufacturing-Priority-Agenda.pdf

[9] Balke,NS.Detecting levelshiftsin timeseries. JournalofBusiness &EconomicStatistics. 1993;11(1):81-92.

[10] HassaniH,MahmoudvandR.MoreonFilteringandForecastingbySSA.InSingular Spectrum Analysis. 2018: 103-115.

[11] HassaniH,ThomakosD.Areviewonsingularspectrumanalysisfor economicand financial time series. Statistics and Its Interface. 2010; 3(3): 377-397.

[12] GolyandinaN,NekrutkinV,Zhiglja vskyAA.Analysisoftimeseriesstructure: SSAand related techniques. CRC Press. 2001 Jan 23.
[13] BansalA,DimriV.Depthdeter minationfromanon-stationarymagneticprofilefor scaling geology. Geophysical Prospecting.
2005;53(3):399-410.

[14] GhilM,AllenM,DettingerM,IdeK, KondrashovD,MannMetal.AdvancedS pectral Methods for Climatic Time Series.
ReviewsofGeophysics. 2002;40(1):31-41.
[15] RekapalliR,TiwariR,Dhanam K,SeshunarayanaT.T-xfrequencyfilteringofhigh resolution seismic reflection data using singular spectral analysis. Journal of Applied Geophysics. 2014; 105: 180-184.
[16] GhilM,VautardR.Interdecadaloscill ationsandthewarmingtrendinglobaltemp erature time series. Nature. 1991; 350(6316): 324-327.
[17] Sanei S, Ghodsi M, Hassani H. An adaptive singular spectrum analysis approach to murmurdetectionfromheartsounds.Me dicalEngineering&Physics.2011;33(3) :362-367.
[18] Rodríguez-AragónL,ZhigljavskyA.Singularspec trumanalysisforimageprocessing. Statistics and Its Interface. 2010; 3(3): 419-426.
[19] deCarvalhoM,RodriguesPC,Ru aA.TrackingtheUSbusinesscyclewit hasingular spectrum analysis. Economics Letters.
2012;114(1): 32–5.
[20] VautardR,YiouP,GhilM.Singular-spectrumanalysis:Atoolkitforshort,noisy chaotic signals. Physica D: Non-linear Phenomena. 1992; 58(1-4): 95-126.
[21] BricenoH,RoccoC,ZioE.Singula rspectrumanalysisforforecastingofel ectricload demand. Chemical Engineering Transactions. 2013; 33: 919-24.
[22] SilvaES,HassaniH.Ontheuseofsi ngularspectrumanalysisforforecastin gUStrade before, during and after the 2008 recession. International Economics. 2015; 141: 34-49.
[23] HassaniH,WebsterA,SilvaES,Hera viS.ForecastingUStouristarrivalsusing optimal singular spectrum analysis. Tourism Management. 2015; 46: 322-35.
[24] Vile JL, Gillard JW, Harper PR, Knight VA. Predicting ambulance demand using singularspectrumanalysis.Journalofthe OperationalResearchSociety.2012;63(1 1):1556- 65.
[25] MenardS,editor.Handbookoflongitu dinalresearch:Design,measurement,anda nalysis. Elsevier; 2007 Oct 11.
[26] AlterO,BrownPO,Botstein D.Singularvaluedecomposition forgenome-wide expression data processing and modeling.
ProceedingsoftheNationalAcademy ofSciences.2000;97(18): 10101-6.
[27] DattaB.Numericalmethodsforlinearc ontrolsystems.AcademicPress;2004.
[28] GhilM,TariccoC.Advancedspectral - analysismethods.InPastandpresentvaria bilityof the solar-terrestrial system: measurement, data analysis and theoretical models. IOS Press. 1997: 137-159.

[29] Silva ES, Hassani H, Heravi S. Modeling European industrial production with multivariatesingularspectrumanalysis:Across-industryanalysis.JournalofForecasting. 2018; 37(3): 371-84.
[30] SyukurA,MarjuniA.StockPriceForecastingUsingUnivariateSingularSpectral Analysis through Hadamard Transform. InternationalJournalofIntelligent EngineeringandSystems.2020;13(2):96-107.
[31] GolyandinaN,KorobeynikovA.Basicsingularspectrumanalysisandforecastingwith R. Computational Statistics & Data Analysis. 2014; 71: 934-54.
[32] BoxGE,TiaoGC.Achangeinlevel ofanon-stationarytimeseries.Biometrika.1965; 52(1/2): 181-92.
[33] Kenya National Bureau of Statistics. Statistical Abstract 2010. [Internet] 2010. [cited 2021August5]Availablefrom:https://www.knbs.or.ke/?wpdmpro=statistical-abstract-2010
[34] KenyaRevenueAuthority.Importing &exporting[Internet].KRA.2020[cited2021July 7]. Available from: https://kra.go.ke/en/business/companies-partnerships/companies- partnerships-pin-taxes/company-partnership-imports-exemptions