The Import Trade Forecasting Model Based on PCA: Evidence from Rwanda

Elie Nzayisenga, Yongzhong Zhu

College of Science, Hohai University, Nanjing, China
Email: esengaa@outlook.com

Abstract

A developing country like Rwanda heavily is keen on international trade for several essential goods in the development of an economy. This study investigated the influence of various factors affecting import trade, and use principal component analysis to determine an empirical model for a comprehensive analysis of the influencing factors of import trade of Rwanda using secondary data over the period from 1980-2017. The PCA model showed that Rwanda's import trade is principally littered with investment fundamental factors, income consumption factors, price factors, inflation factors, and savings factors and the empirical results showed that Rwanda's import trade is negatively correlated with the investment fundamental and savings factors, the income consumption factors, price factor, and the inflation are positively correlated and therefore the forecast for the period 2018-2025 revealed that the import trade of Rwanda may experience an increase. The implication is that unstable price and currency depreciation cause high income consumption and increased import trade volume. The study advises policy makers on international trade first to pay attention to the accumulation of investment and savings checking if providing support for import trade control and enhance economic security. Second, stabilize the price and manage to keep inflation low and stable. Third, better focus on improving domestic production by not permitting Rwandan currency (Frw) to lose the worth, thus directly forming the necessity for foreign merchandise for investment purposes to increase the level of production exportation, which might have a giant positive impact on saving culture linked to economic growth.

Keywords

Principal Component Analysis, Import Trade, Influencing Factors, Policymakers

1. Introduction

International trade has existed for a protracted time [1]. Not all countries have

enough resources and skills required to produce certain goods and services. Nevertheless, countries impose trade barriers in order to protect their domestic industries. However, trade has accrued vastly within the past few hundred years and contains a major impact on the economies of many countries [2].

An inland country with 26.338 km^2 and a densely packed population of about 12.2 million people, Rwanda has reached 6.1% of economic growth rate in 2017. A touch economic transformation after Genocide of Tutsi 1994 [3], given the country's move towards vision 2020, attracted international investors and explicit challenges in leveraging FDI for development as a result of its economic structure [4] and commodity-producing countries resorted to importation of more intermediate and capital goods from abroad for its import substitution and industrialization programs as shown in Figure 1 to support and sustain economic growth [5] but costs of trading are higher due to long land transport routes, lack of proper infrastructure and the high import and export freight service costs within the region. In light-weight of those circumstances, it has become essential to find out which factors explaining the abrupt variations and continuous growth in imports for Rwanda.

Imports play a crucial role in the investment atmosphere and therefore industrial development and economic progress of a country. To check the foreign trade regime of a country, a stronger image might be obtained if imports are studied at some length.

Principal component analysis (PCA) is often used for analyzing data in the most diverse areas. Although macroeconomic variables have long been studied, many economists and researchers have often neglected the preliminary study of the variables regarding their similarities and differences [6]. In this study of analyzing factors influencing import trade of Rwanda and its forecasting, we consider PCA because of its efficiency in removing redundancy where manipulation of large numbers of attributes is necessary and capability of extracting interpretable information in case. There are an outsized variety of correlated variables that usually contain loads of information redundancy.
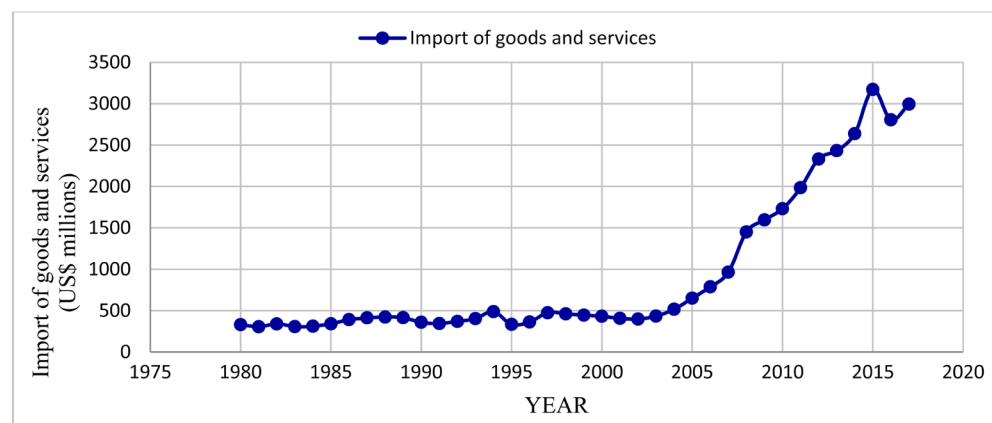
Source: Data are in US$ million and taken from world development index (WDI).

Figure 1. Trend in import of goods and services from 1980-2017.

PCA is used to cut back this redundancy, which ends within the reduction of highly correlated data into a small number of uncorrelated principal components that truly account for most of the variance within the highly correlated data [7].

Despite the apparent simplicity of the techniques, abundant analysis remains being drained the overall space of PCA, and it is terribly wide used and also the explosion of latest applications and more theoretical developments have occurred. This growth reflects the general expansion of the statistical literature, however as PCA requires hefty computing power, the expansion of its use coincided with the widespread introduction of electronic computers. Most of the applications of PCA in Economy for validation as a method to explain the socio-economic status differentiation among population and evaluating the financial development of nations or entities employing a combination of features ought to be thought considered once generating and interpreting results [8].

This study will undoubtedlycontribute to the prevailing empirical literature on the issue of factors influencing import trade using the Rwandan situation and be a profit to the importers and exporters (whether companies or individuals) by letting them perceive how their activities contribute to the development of Rwanda.

This study is worked out based on principal component analysis, which focused on removing redundancy and compress factors influencing import trade of Rwanda into few components to be used in principal component regression analysis. The method and data part is focused on data source, variables selection and model estimation, while the empirical analysis show the associated results as well discussion of the founded results. Then conclusion summarizes the findings and recommendation part shows opinions supported by the research's findings for policy implications.

2. Data and Methods

Principal component analysis is a popular data-processing and dimension-reduction technique with optimal properties as the linear combinations of the original variables such that the derived variables capture the maximum variability guaranteeing minimal information loss and principal components are uncorrelated where we can interpret each one independently [9].

Many economic variables have the property that they are correlated with each other. Given the natural links between almost all facets of economic activity within any given economy, the multicollinearity is severe. The PCA objectives are to extract an important part of the information from the data set, reducing the size of data with minimal damage to data and information. This is achieved by finding a new set of independent (uncorrelated) variables called principal components which are obtained as a linear combination of the original variables.

The calculation of PCs means the computation of eigenvalues and eigenvectors for a positive-semidefinite symmetric matrix. The first PC has the largest proportion of variance of the data, and the second component has the second largest proportion of variance and is orthogonal to the first principal component. Remaining PCs represents the remaining variance in descending order,

and each PC is orthogonal to its predecessors. when computing the PCs, the primary first PCs that represents the large part of variation are designated to be used in further analysis.

2.1. Data

This study has tried to incorporate all factors in the import function that can potentially play a meaningful role in the import trade model determination for Rwanda. The data are obtained from WDI and UNCTAD officially recognized international sources and the selected variables are summarized in Table 1. In view of the availability of data, selected was from 1980 to 2017 and form a total of 456 observations.

The motivation for employing the principal component to these indicators is, most of these variables are often speculated to be highly related to one another and hence the precision of estimation may be questionable if we will in the near future want to determine the impact of these indicators on a particular event. The outputs are clearly learned from R statistical software package 4.0.2, IBM SPSS Statistics 20 and Microsoft excel 2016.

2.2. Model Estimation

Principal component analysis is a popular data-processing and dimension-reduction technique with optimal properties as the linear combinations of the original variables such that the derived variables capture the maximum variability guaranteeing minimal information loss and principal components are uncorrelated where we can interpret each one independently [9].

The data consist:

Table 1. Variables selected for building the import trade model.

Variables	Names
Y	Import volume index
X_1	Foreign direct investment, net inflows (% of GDP)
X_2	Gross capital formation (% of GDP)
X_3	Official exchange rate (LCU per US$, period average)
X_4	Trade openness
X_5	External debt stocks (% of exports of goods, services and primary income)
X_6	Final consumption expenditure (% of GDP)
X_7	Relative import prices
X_8	GDP growth (annual %)
X_9	Inflation GDP deflator (annual %)
X_{10}	Population growth (annual %)
X_{11}	Savings [Total reserves (% of total external debt)]

Source: author's own.

Suppose $X = (X_1, X_2, \cdots, X_n)^T = (\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_p)$ is the design matrix of size, $n \times p$ columns represent p explanatory variables and rows represent n samples.

$$X = \left[x_{ij} \right]_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{2.1}$$

We assume that all columns are centered and $\mu = E(X), \Sigma = VAR(X)$. Principal components are derived from the linear transformation $Z = XV$, where $V \in \mathbb{R}^{p \times r}$ with $r = rank(X)$ is the orthogonal matrix of principal component loadings. It is worth mentioning that the PC transformation preserves the total variance (V_T) from the initial space:

$$V_T = \sum_{i=1}^{n} Var(x_i) = \sum_{i=1}^{n} \lambda_i \Leftrightarrow tr(\Sigma) = tr(\Lambda) \tag{2.2}$$

where Λ is the covariance matrix of the PCs is a diagonal matrix

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \tag{2.3}$$

PCA also results in the preservation of generalized variance (V_G):

$$V_G = |\Sigma| = |\Lambda| \tag{2.4}$$

$\Lambda = diag(\lambda_1, \cdots, \lambda_p)$ contains the (non-negative) eigenvalues, the principal components $Z_1, Z_2, \cdots, Z_p$ should therefore capture much information as possible.

The proportion of the total variance explained by the pth principal component can be written as:

$$W = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, \quad (i = 1, 2, \cdots, p) \tag{2.5}$$

where λ_i, is the eigenvalue of the pth principal component. For instance, the eigenvector with the highest eigenvalue is the principal component of the data set. If the first principal components can explain most of the variations in the sample covariance, the p variables can replace the original k variables with little loss of information.

The cumulative percentage of eigenvalues gives the percent variability of the data set

$$V = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i}, \quad (i = 1, 2, \cdots, p) \tag{2.6}$$

where V is accumulated variance contribution rate and the values of the eigenvectors (original dimension) of each PC (which vary from −1 to +1) can be in-

terpreted as an index of the combined action or contrast of the original dimensions. Thus, principal components for the ith sample is given by $Z_i = V^T X_i$ for $i = 1, 2, \cdots, n$, where Z_i is the ith row of $Z = (Z_1, Z_2, \cdots, Z_n)^T = (\tilde{Z}_1, \tilde{Z}_2, \cdots, \tilde{Z}_r) \in \mathbb{R}^{n \times r}$.

The principal component loadings:

$$a_{ij} = \sqrt{\lambda_i e_{ij}}, \ (i, j = 1, 2, \cdots, p) \tag{2.7}$$

Such that $Z = a^T X$. The first principal component is defined as the variable, $Z_1 \equiv a_1^T X = \sum_{i=1}^{p} a_{1i} x_i$, $Z_1 = a_{11} x_1 + a_{12} x_2 + \cdots + a_{1p} x_p$ which has the maximum variance with the constraint that $a^T a = 1$.

$$\begin{cases} Z_1 = a_1^T X \\ \quad \vdots \quad \quad \text{var}(z_i) = a_i^T \Sigma a_i, \ i = 1, 2, \cdots, 11 \\ Z_i = a_i^T X \end{cases} \tag{2.8}$$

and have $COV(Z_i, Z_j) = a_i^T \Sigma a_j^T, i, j = 1, 2, \cdots, 11; i \neq j$.

In this representation, the columns of Z are mutually orthogonal, but not normalized. We will frequently re-express $Z = UD$ where U is orthogonal matrix of scaled principal component scores and D is diagonal matrix with the square-root of variance of principal component score. Thus, U and V are matrices of the left and right singular vectors, and $D = diag(d_1, \cdots, d_r)$ is a diagonal matrix of ordered singular values of X.

With the response variable $Y = (y_1, y_2, \cdots, y_n)^T$, the linear regression model is typically assumed to be:

$$Y = \alpha + X\beta + \varepsilon \tag{2.9}$$

where α is the intercept, $\beta = (\beta_1, \cdots, \beta_p)^T$ is the vector of regression coefficients, and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$ is a vector of random error of mean 0 and variance $\sigma^2 I_n$, i.e. $\varepsilon \sim N_n(0, \sigma^2 I_n)$. X is centred, OLS estimator of α is given as $\bar{y}$; therefore, we remove the intercept in the regression model with the centred response variable. OLS estimate of β is obtained by minimizing the sum of squared errors over β:

$$\hat{\beta}^{OLS} = \arg\min_{\beta} \|y - X\beta\|_2^2 = \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} \beta_j \tilde{X}_j \right\|_2^2 \tag{2.10}$$

Principal component analysis belongs in the general framework of multivariate statistical analysis techniques. Principal component regression is a method to perform data reduction in the model as well as to solve problems of dependencies among variables (multicollinearity) [10]. PCR is performed under the least square's errors criterion using principal component scores as explanatory variables rather than the original explanatory variables:

$$\hat{\gamma} = \arg\min_{\gamma} \|y - Z\gamma\|_2^2 = \arg\min_{\gamma} \left\| y - \sum_{j=1}^{r} \gamma_j \tilde{Z}_j \right\|_2^2 \tag{2.11}$$

where $\gamma = (\gamma_1, \cdots, \gamma_r)^T$. Since $X\hat{\beta} = ZV^T\hat{\gamma}$, we can obtain an OLS estimate of β

with the relationship of $\hat{\beta}^{OLS} = V\hat{\gamma}$. Let $Z_k = \left(\tilde{z}_1, \cdots, \tilde{z}_k\right)$ where $\tilde{z}_j$ is the jth column of Z, *i.e.*, the jth principal component score of X. The common practice of PCR is performed with the first k principal component scores:

$$\hat{\gamma}_k = \arg\min_{\gamma_k} \left\| y - Z_k \gamma_k \right\|_2^2 = \arg\min_{\gamma_k} \left\| y - \sum \gamma_j \tilde{Z}_j \right\|_2^2 \qquad (2.12)$$

where $\gamma = \left(\gamma_1, \cdots, \gamma_k\right)^{\mathrm{T}}$, PCR estimate of β is given as $\hat{\beta}^{PCR,k} = V\hat{\gamma}_k$.

The number of principal components used in PCR, k, is typically chosen by considering the proportion of variance explained and prediction performance using cross validation [11].

3. Findings

3.1. KMO Test of Data and Spherical Disturbance Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy test if the sample size used for the study is adequate for principal component analysis and Bartlett's test of homogeneity of variances test the hypothesis that correlations between variables are greater than would be expected by chance. It was used to compares the correlation matrix with a matrix of zero correlations usually known as an identity matrix, which consists of all zeros except the 1's along the leading diagonal. The results are presented in Table 2.

Due to high MSA, the overall KMO value of 0.85 indicates that the sample size used for the study is adequate and that principal component analysis of the data is more appropriate. It can be seen that the p-value is approximately equal to 0.000 (<2.22e−16), indicating that the data is from a normal state and there is a strong correlation between the variable series.

Now face the problem of information redundancy due to powerful correlations between the variables. It is important to understand how these could adequately impact on a particular incidence without necessarily compromising estimation precision. The factors influencing the import trade of Rwanda as shown in the table are highly correlated, meaning that they vary together.

Table 2. Kaiser-Meyer-Olkin Statistics and Bartlett's Test of Sphericity results.

Kaiser-Meyer-Olkin Statistics											
KMO-Criterion: = 0.85											
Measures of Sampling Adequacy (MSA):											
Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
0.85	0.84	0.88	0.89	0.91	0.89	0.92	0.79	0.82	0.90	0.86	0.70
Bartlett's Test of Sphericity											
Approx. Chi-square						1530.18					
Df						66					
p-value						<2.22e−16					

Source: Author's own.

This redundancy in the measurements allows us to build a PCA model that will retain most of the information in a few principal components.

3.2. Total Variance Explained

The total variance explained which indicates how much of the variability in the data has been explained by the components shown in Table 3 and it is very important to distinguish the number of components (dimension) that are to be kept in the model.

It is necessary to decide on the number of components that have any practical significance. A simpler but arbitrary rule of thumb, which has proved to be useful in practice, the decision is based on the number of components with a standard deviation greater than or equal to one as having any practical significance. In this study, the first four components, accounting for about 86.3% of the total variability would be regarded as being of practical significance, although the possible interpretation of the next component, bringing the total variability to about 92.9%, would also be considered.

3.3. PCA Loadings and Interpretation

PCA models are typically very easy to interpret and when interrogated can provide fault diagnosis [12]. A threshold of 0.60 was used for identifying a reliable component and in this study indicated that items that load less than 0.50 are not accepted, all selected components are having significant loadings above 0.60.

PCA gives new indicators which are linear combinations of the original ones, thus the new indicators combines similar old indicators through their shared properties, Five main components were extracted to represent the 11 variables, the first PC is powerfully correlated with four of the original variables (Foreign

Table 3. Eigenvalues of the correlation matrix of influencing factors of import trade.

| Component	Eigenvalue	Percentage of variability	
		Component	Cumulative
PC1	4.475	40.679	40.679
PC2	2.285	20.776	61.455
PC3	1.486	13.509	74.964
PC4	1.253	11.389	86.353
PC5	0.725	6.590	92.943
PC6	0.353	3.205	96.148
PC7	0.188	1.710	97.858
PC8	0.112	1.022	98.881
PC9	0.080	0.730	99.611
PC10	0.029	0.266	99.877
PC11	0.014	0.123	100.000

direct investment, gross capital formation, trade openness and exchange rate) and recovers 40.679% of the variability, three (3) variables correlate well with the second component, one (1) variable correlate well with the third component, two (2) variables correlate well with the fourth component, one (1) variable correlate well with the fifth component.

As can be seen from the principal component load matrix in Table 4 and Figure 2, eleven original explanatory variables reflect the basic economic situation of a country.

Table 4. Rotated principal component load matrix table.

	Component				
	PC1	PC2	PC3	PC4	PC5
Foreign direct investment	0.789	−0.458	−0.006	−0.105	0.264
Gross capital formation	0.836	−0.481	0.095	−0.034	0.107
Official exchange rate	0.924	0.167	0.260	−0.042	0.025
Trade openness	0.736	0.211	−0.489	0.043	0.326
External debt stock	−0.254	0.801	−0.225	0.179	0−.403
Final consumption expenditure	−0.141	0.740	−0.584	0.168	−0.038
Relative import prices	0.049	0.966	0.077	0.022	−0.109
GDP growth	0.096	−0.079	0.946	−0.002	0.075
Inflation	−0.089	0.251	0.185	0.892	0.119
Population growth	−0.017	0.091	0.430	−0.781	0.286
Savings	0.270	−0.320	0.066	−0.043	0.857

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Source: Author's own.
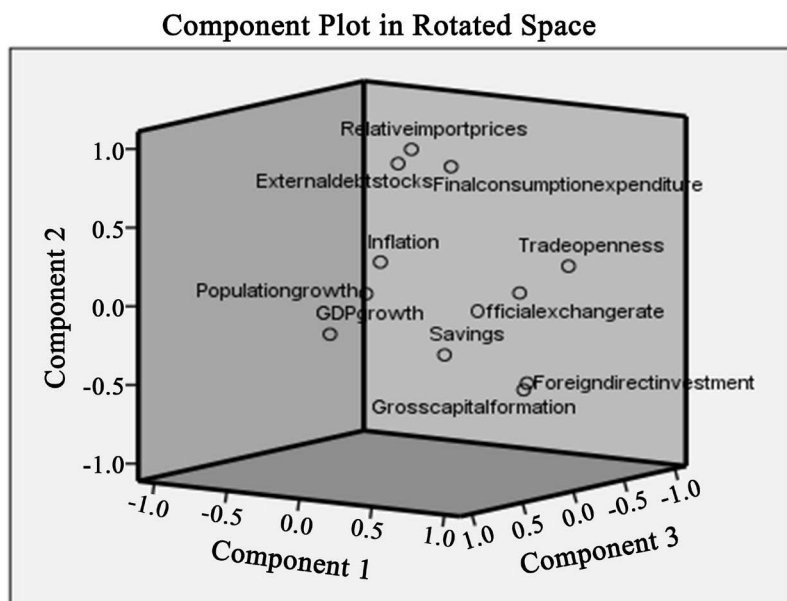
Figure 2. Component plot in rotated space. Source: Author's own.

Having determined the number of principal components retained, we are now concerned with the interpretation of the PCs:

The first component is positively correlated with Foreign direct investment, gross capital formation, trade openness and exchange rate. This correlation suggests the four variables vary together and when the others decrease as well and could be considered as primarily a measure of investment. The second component is most correlated with external debts, final consumption expenditure and relative price import, both in a positive direction. As price as signals of the competitiveness of a country's exchange power vis-à-vis the rest of the world in market reflects the income consumption information. The third component is positively correlated with gross domestic production and not much else, this component could be viewed as a monetary measure of the market value of all the final goods and services produced in a specific time period reflecting the price information. The fourth principal component is inflation and population growth, the component most correlated with inflation at 0.892 and could be considered as primarily a measure of inflation information and the fifth components of total reserves reflects saving information.

3.4. Principal Components Regression Results

Principal component regression (PCR), is a technique that combines principal component analysis with inverse least-squares regression [13]. Due to these characteristics PCA and its regression (PCR) were selected to be used to establish the relationship between principal components retained and import volume index for prediction. The contribution of PCR to model improvements mainly derives from the power of the PCA to capture the original variability of the dataset in a small set of PCs while these PCs are by default uncorrelated.

Using 5 named principal components as explanatory variables (set with the import amount (Y) as the explanatory variable component regression to further analyze the relationship between import trade and principal components relationship.

Table 5. Principal components regression results.

Coefficients:				
	Estimate	Std. Error	t-value	Pr (>|t|)
(Intercept)	250.74	9.84	25.49	<2e−16***
β_1	−79.56	4.71	−16.88	<2e−16***
β_2	74.46	6.60	11.29	1.1e−12***
β_3	13.34	8.18	1.63	0.11
β_4	13.42	8.91	1.51	0.14
β_5	−99.50	11.71	−8.50	1.0e−09***

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1. Residual standard error: 60.6 on 32 degrees of freedom. Multiple R-squared: 0.939, Adjusted R-squared: 0.929, F-statistic: 97.9 on 5 and 32 DF, p-value: <2e−16. Source: Author's own.

Table 5 shows that both the regression coefficient and the regression equation pass the test and at least 3 are highly significant; in the constant term test, the *p*-value is less than zero and the effect is remarkable.

The function reflects the relationship between the import volume index and each principal component. The estimated PCR that fits the data gathered is given as:

$$Y = 250.74 - 79.56Z_1 + 74.46Z_2 + 13.34Z_3 + 13.428Z_4 - 99.5Z_5$$

The import trade has a significant negative relationship with investment fundamentals and saving factors, positive significantly related with income consumption factors and positive non-significant with price and inflation factors. The F-statistics was statistically significant at a 5% significance level (F = 97.9, *p*-value < 2e−16). Also, the R-square was approximately 92.9% to show how much the components can be explained on the dependent variable.

In Figure 3, Residuals versus fitted plot showed that that the model is not linear, the normal Q-Q plot showed that model is skewness on the right than would be expected theoretically. The analysis of the residuals vs leverage plot to detect outliers in the model showed how the spread of standardized residuals changes as the leverage and there is no point outside the red line. The scale location plot here showed that the average magnitude of the standardized residuals isn't changing much as a function of the fitted values, our data generally has uniform variance at the ends of our predictor range and is somewhat heteroscedastic in the middle of the range.
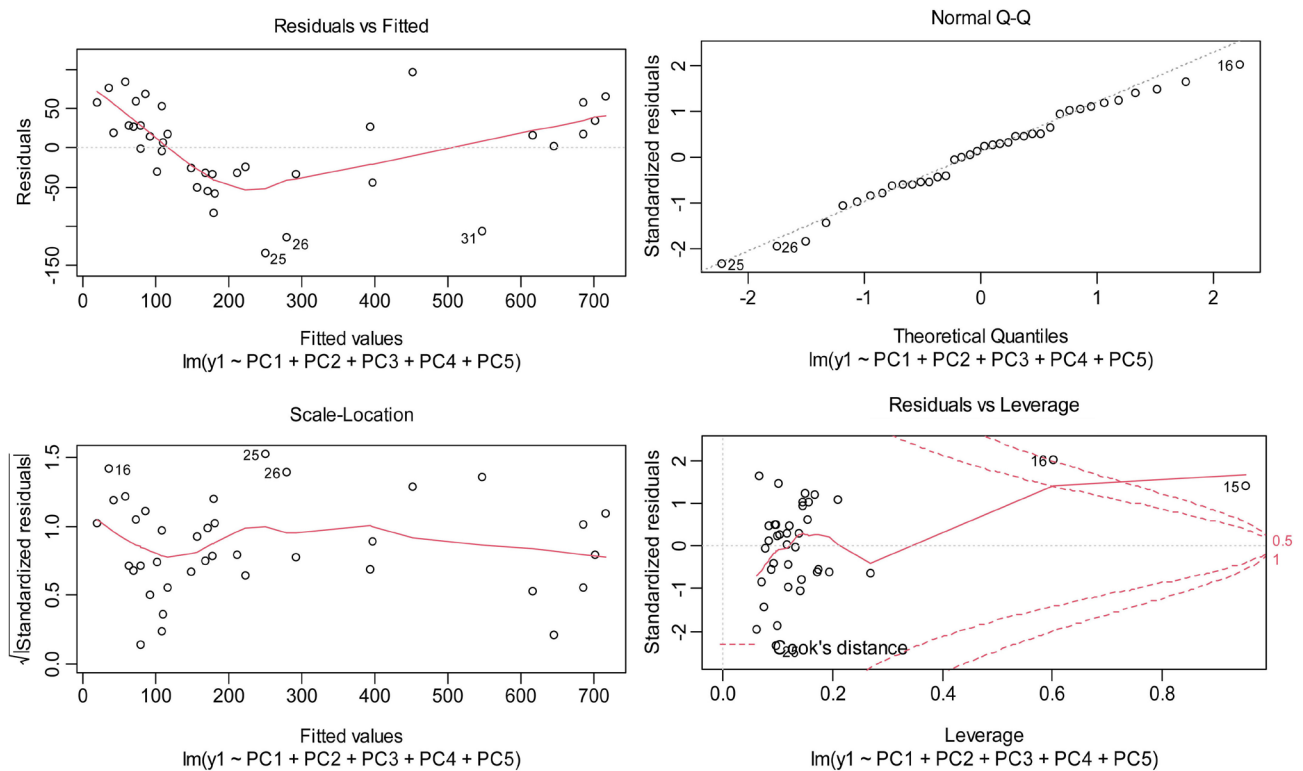
**Figure 3.** Residual plot. Source: Author's own.

3.5. Data Partition

Data splitting is the act of partitioning available data into two portions. Usually, for cross-validation purposes [14] a split data analysis is preferred to full data analysis for prediction. Data splitting divides the dataset into two sets: Training set and Testing set. the training set is used to build and train the model and tested on the testing set for accuracy and well performance [15]. The objective is to have the model perform on any data with the highest accuracy, the data are partitioned into 1980-2004 for the training set and from 2005-2017 for the testing set. These data are used for actual prediction applications.

11 explanatory variables of the training set are compressed into 5 components and their scores are regressed on import volume index; the prediction shows the accuracy of 0.89. The same technic is used to the testing set and the accuracy was 0.90. The model has been trained well and tested thoroughly, the evaluation of the performance was based on MIN-MAX Accuracy after PCA and we combined the database created: the database of actuals and predicted training set and database of actuals and predicted test set into one and form **Figure 4**; both lines seem to be close to each other. Therefore, we can say that the principal component analysis-based modelling is a good one for predicting the import trade of Rwanda.

3.6. Import Trade Forecasting

Forecasting future observations is the most important task of time series analysis when there are many predictors. The data followed an approximate factor model, here the predictors were summarized by a small number of components estimated using principal components [16]. Holt-Winters smoothing is a widely used tool for forecasting business data that contain seasonality, changing trends, and seasonal correlation [17]. The Holt-Winters method is incredibly important, though, as it is one of the reasons why it is so attractive and easy to use [18], my main interest is the forecast performance of the smoothing methods. Therefore,
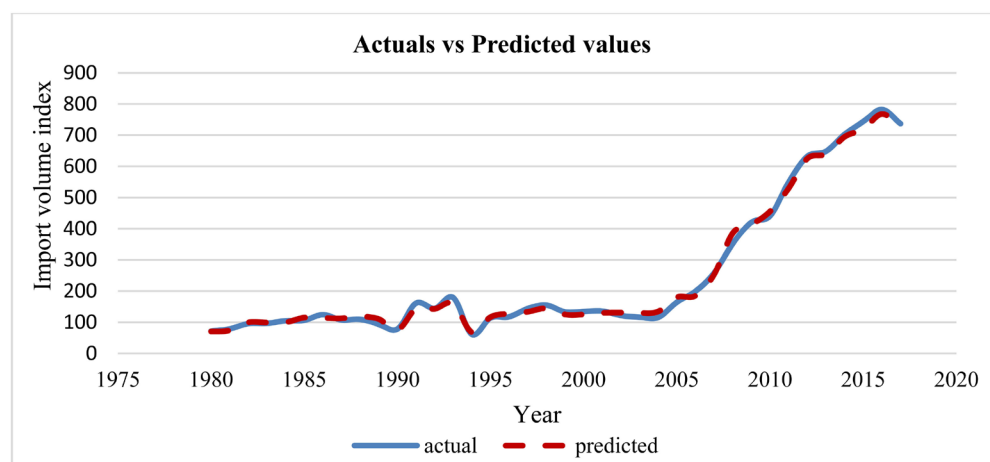
Figure 4. Graph of actuals and predicted values. Source: Author's own.
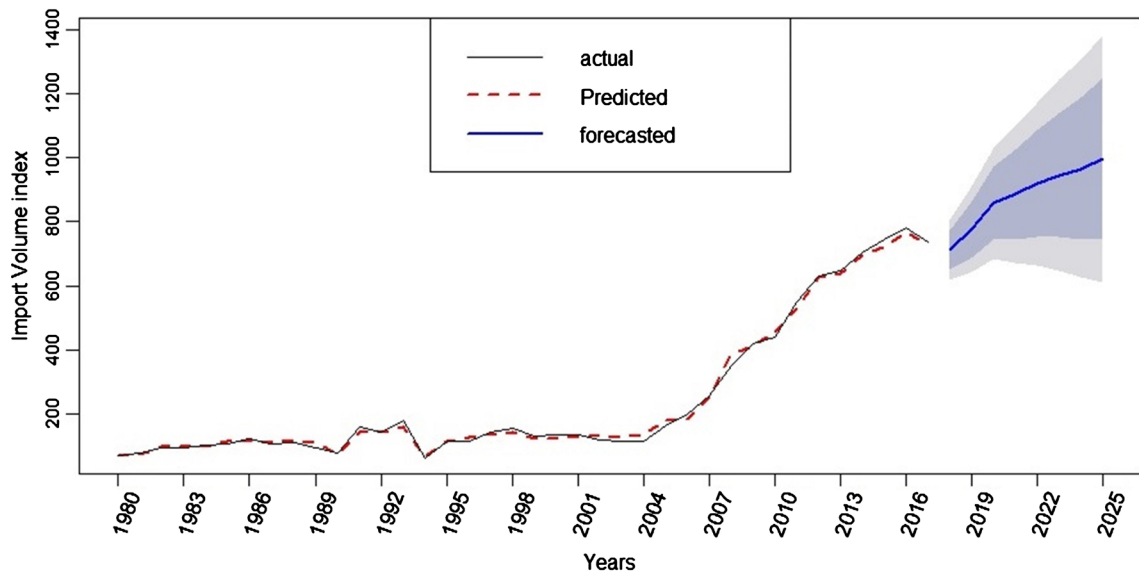
Figure 5. Graph of Rwanda's import trade forecasting to 2025. Source: Author's own.

we select holt winter's smoothing due to its prediction abilities and that it can adapt to the changing trend.

Figure 5 shows the forecast values for the import volume index of Rwanda for the following 8 years. This figure also includes the low and high values for 80% and 95% confidence intervals.

The black line represents the actual Import volume index from 1980-2017, the red line represents the predicted values of components influencing import trade from 1980-2017 and the blue line represents the forecast for the eight years 2018-2025. The shaded region represents the confidence intervals values. Figure showed that the import trade of Rwanda for the following 8 years will experience an increase.

This result is in line with the annual report given by the world bank database, which gave an import volume index of 776,339 for 2018, it is included in the confidence interval.

4. Conclusions and Recommendation for Policy Implication

4.1. Conclusions

Based on the findings above, the eleven macroeconomic indicators showed fairly a degree of unpardonable redundancy in the correlation matrix. Preliminary to PCA, we found that five PCs will suffice. From this, the PCA was adopted to identify the variables that are likely to be classified together as one principal component. Five PCs roughly explain 92.4% of the total variation in the data. We found that four of those indicators including foreign direct investment, gross capital formation, trade openness, and exchange rate showed a very high likeness and these were classified into one factor referred to as the investment funda-mental factor. External debts, final consumption expenditure, and relative price import exhibited great similarities and were classified into another factor re-

ferred to as the income consumption factor. The third principal component is the gross domestic product (GDP) as a monetary measure of the market value of all the final goods and services produced in a specific time period can be named as the price factor. The fourth principal component is inflation and population growth reflects the inflation factor but total reserves on its own uniqueness from the others were classified as the saving factor.

After the extraction of the five main components, the eigenvectors for the five components were used as regressors for the regression analysis and the results show that there is at least a significant component in the model.

The empirical model showed that import volume has a positive relationship with income consumption, pricing and inflation factors and negative relationship with investment fundamental and saving factors, that is certainly and generally accepted that the import trade increased by price instability, high consumption and current account deficits lead to currency problems (high inflation).

Investment fundamentals and savings factors had a big negative impact on import volume as an increase in domestic investment repatriated a huge amount of 1673 million USD from abroad [19] in 2017 and gross domestic savings (% of GDP) in Rwanda was reported at 7.5955% in 2018, according to the World Bank collection of development indicators thus reducing the amount of money to borrow and buying in foreign currencies to pay for consumption.

An increase in the price of domestic products compared to imported products increased price instability and encouraged diversification of imported products and spending more income to avoid future price increases and involves an increase of import volume. Consumption increases when disposable income increases and countries consume more imports as the income of their population rises, the reason for the positive relationship between income consumption and import volume.

The forecasting model improves the accuracy of prediction and reduces the error between the actual value and the forecast value in import trade prediction. Using PCA method, the best feature subset for forecasting can be found and the impact of redundant information between factors influencing import trade can be reduced. As a result, the accuracy of forecasting is improved. Finally, the application of PCA served as tools for effectively reducing and grouping variables into fewer components with little loss of valuable information. This provides the essence of the principal component analysis.

4.2. Recommendation for Policy Implication

Rwanda is currently establishing a competitive knowledge-based economy and promoting industrial enterprise development and also the incentives of investment targeted at promoting exports and shifting from an agriculture-based economy to high-value goods and services [20]. Promoting stable and sustainable development of import trade is conducive to improving Rwanda's current foreign trade pattern that relies heavily on tea and coffee exportation, and con-

tributes to the transformation and upgrading of the economic structure and the overall welfare improvement.

Therefore, Rwanda's economic policymakers have to be compelled to take into consideration the following three points to manage import trade promotion:

First, as Rwanda is fast developing economically, technologically and mentally and there is an enormous need for innovative products, leading to Rwanda's economy higher quality, higher efficiency, fairer and more sustainable that why Rwanda economic policymaker should control the accumulation of the investment by encouraging higher private investments and savings to enrich public funding, industrial policy and human capital thus providing support import trade control and economic security.

Second, prices change due to supply and demand, has an affect import trade positively, to attain sustainable economic growth coupled with price stability. Policy authorities should manage to keep inflation low and stable with negotiating market access through new bilateral trade agreement and transport agreement and conjointly establishing regular monitoring of the price on the market.

Third, the rise in the exchange rate indicates the national currency (Frw) depreciation also will have a negative impact on imports, independency, investment growth as well as economic resilience. With speedy economic development and growth of import trade in Rwanda coming from coffee and tea base, capital accumulation is the main driving force for economic growth. Therefore, the growth of investment and savings can be driven by weakening import trade which will serve as a catalyst to attract more external financiers and investors as it reduced the value of the international flow of capital, thus directly forming the need for imported products for investment purposes to increase the level of product exportation this would have a giant positive impact on saving culture linked to economic growth.

Authors' Note

This manuscript is the authors' original work, has not been published and is not under consideration for publication elsewhere.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Dubravska, M. and Sira, E. (2015) The Analysis of the Factors Influencing the International Trade of the Slovak Republic. *Procedia Economics and Finance*, **23**, 1210-1216. https://doi.org/10.1016/S2212-5671(15)00569-9

[2] Nadeem, M., Azam, M. and Islam, R. (2012) An Investigation of the Various Factors Influence. *Global Journal of Management and Business Research*, **12**, 56-62.

[3] Khan, I., Mughal, K., Zia, F. and Usman, F. (2016) Do the GDP, Budget Deficit, External Debts, Exports and Imports affect Each Other for Pakistan: An ARDL Approach. *Journal of Economics and Finance*, **7**, 18-30.

[4] United Nations Conferences on Trade and Development (UNCTAD) (2006) Investment Policy Review Rwanda. New York and Geneva.

[5] African Development Bank Group (AfDB) (2019) East Africa Economic Outlook 2019. Macroeconomic Developments and Prospects and Political Economy of Regional Integration. Abidjan.

[6] Armeanu, D. and Lache, L. (2008) Application of the Model of Principal Components Analysis on Romanian Insurance Market. *Theoretical and Applied Economics*, **6**, 11-20.

[7] Waqar, M., Dawood, H., Shahnawaz, M.B., Ghazanfar, M.A. and Guo, P. (2017) Prediction of Stock Market by Principal Component Analysis. 13*th International Conference on Computational Intelligence and Security*, Hong Kong, 15-18 December 2017, 599-602. https://doi.org/10.1109/CIS.2017.00139

[8] Vyas, S. and Kumaranayake, L. (2006) Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis. *Health Policy Plan*, **21**, 459-468. https://doi.org/10.1093/heapol/czl029

[9] Zou, H., Hastie, T. and Tibshirani, R. (2006) Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, **15**, 265-286. https://doi.org/10.1198/106186006X113430

[10] Dadousis, C. (2012) The Use of Principal Component Analysis for Predicting Genomic Breeding Values. Wageningen University, Wageningen.

[11] Lee, H., Park, Y.M. and Lee, S. (2015) Principal Component Regression by Principal Component Selection. *Communications for Statistical Applications and Methods*, **22**, 173-180. https://doi.org/10.5351/CSAM.2015.22.2.173

[12] Gallagher, N.B., Wise, B.M., Butler, S.W., White, D.D. and Barna, G.G. (1997) Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness through Model Updating. *IFAC Proceedings Volumes*, **30**, 79-84. https://doi.org/10.1016/S1474-6670(17)43143-0

[13] Merz, C.J. and Pazzani, M.J. (1999) Principal Components Approach to Combining Regression Estimates. *Machine Learning*, **36**, 9-32. https://doi.org/10.1023/A:1007507221352

[14] Picard, R.R. and Berk, K.N. (1990) Data Splitting. *American Statistician*, **44**, 140-147. https://doi.org/10.1080/00031305.1990.10475704

[15] Faraway, J.J. (2016) Does Data Splitting Improve Prediction? *Statistics and Computing*, **26**, 49-60. https://doi.org/10.1007/s11222-014-9522-9

[16] Stock, J.H. and Watson, M.W. (2002) High-Resolution Analog-to-Digital Converter High Resolution: Flexible Digital Filter: Analog Supply. *Journal of the American Statistical Association*, **97**, 1167-1179.

[17] Gelper, S., Fried, R. and Croux, C. (2010) Robust Forecasting with Exponential and Holt-Winters Smoothing. *Journal of Forecasting*, **29**, 285-300. https://doi.org/10.1002/for.1125

[18] Goodwin, P. (2010) The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong. *Foresight: The International Journal of Applied Forecasting, International Institute of Forecasters*, No. 19, 30-33.

[19] National Institute of Statistics of Rwanda (NISR) (2019) Rwanda Statistical Year-Book 2019, December 2019.

[20] Newfarmer, R. and Twum, A. (2018) Promoting Industrialisation in Rwanda: Five Ideas and Twelve Policy Actions.