



FORECASTING THE GROWTH OF IMPORTS IN KENYA USING ECONOMETRIC MODELS

Eric Ondimu Monayo, Administrative Assistant, Kisii University, Kitale Campus

Alex K. Matiy, Postgraduate Student, Moi University

Edwin Omurumba Mkabane, Postgraduate Student, Moi University

Jefferson Magomere, Postgraduate Student, Moi University

Abstract: The purpose of the study was to forecast the growth of imports in Kenya using the ARIMA model. Box-Jenkins methodology to forecasting was used. Imports data for the period 1960-2015 was obtained from the Trade Map. Data analysis entailed diagnostic tests and fitting the right model for forecasting. Using the proposed model, ARIMA (13, 1, 13), the results of forecasting showed that the growth of imports in Kenya have fairly increasing trend over the nine years forecasted. The study recommends that the government of Kenya should evaluate its import composition. Kenya should work on importing raw materials and unfinished goods so as to assemble them locally to cut costs.

Keywords: Forecasting, Demand, Imports, Econometric models, Kenya.

I. INTRODUCTION

It is generally accepted that developing economies require increasing quantities of certain imports they cannot produce themselves efficiently, especially capital goods, certain intermediate goods and many raw materials. Vigorous import substitution only tempers the economy, but does not eliminate the growing demand for imports. The importance of foreign trade in the development process has been stressed in the two-gap model developed by McKinnon (1964). Imports are a key part of international trade and the import of capital goods in particular is vital to economic growth. Imported capital goods directly affect investment which, in turn, constitutes the engine of economic growth. For Kenya, such goods include heavy machinery and equipment, intermediate inputs and other raw materials, such as crude oil. Initially, most developing countries could place orders with suppliers for any quantity or value of imports. By the early 1970s, they began experiencing chronic foreign exchange problems, which exemplified the looming economic crisis (Mwase, 1990). Thus, in the previous decades, the capacity to import for some developing countries, including Kenya, had declined or stagnated while import demand continued to grow.



Robert (2005) explained that a current account deficit implies an excess of imports over exports of goods, services, investment income and unilateral transfers. This leads to an increase in net foreign claims upon the importing nation. The importing nation becomes a net demander of funds from abroad, the demand being met through borrowing from other nations or liquidating foreign assets. The result is a worsening of the importing nation's net foreign investment position. According to Osoro (2012), increase in the value of imports in most developing economies is largely due to the increase in prices of Petroleum; oil lubricants, fertilizers, and food grains.

The goals of Vision 2030 is to improve manufacturing by reducing imports in key local industries, growing market share in regional market and attracting at least 10 large strategic investors in key agro-processing industries. Kenya is also aiming at strategically increasing the level of value addition in niche exports by additional processing of local agricultural products (Republic of Kenya, 2007). The main goal of Vision 2030 is to promote exports and reduce imports by promoting agro processing.

Despite the expected strong growth in exports of goods and non-factor services, the external current account deficit widened gradually to 6.7 per cent of GDP in 2009, reflecting the strong increase in imports associated with increased foreign direct investment and higher disbursement of long term capital for investment spending. The demand for imports in an economy is a crucial macroeconomic relationship with significant implications for the design and conduct of economic policy. In view of the importance of imports to the growth process of economies, especially in developing countries, a number of empirical studies on import demand elasticity have been conducted especially in Asia and Latin America and a few in African countries. What is evident, no single study has forecast the growth of imports in Kenya. This gap was therefore addressed by the current study.

II. METHODOLOGY

2.1 Data

The study researches on imports yearly data, which have been collected for the period 1960-2015, sourced from the Trade map. EViews and Excel are the main statistical software for analysis and estimation employed in this study.



2.2 ARIMA modeling

The basic idea underlying the Box-Jenkins methodology to forecasting is to analyze the probabilistic, or stochastic, properties of economic time series on their own under the philosophy “let the data speak for themselves”. Unlike traditional regression models, in which the dependent variable Y_t is explained by k explanatory variables $X_1, X_2, X_3, \dots, X_k$, the BJ time series models allow Y_t to be explained by the past, or lagged, values of Y_t itself and the current and lagged values of u_t , which is an uncorrelated random error term with zero mean and constant variance that is, a white noise error term.

The Box-Jenkins ARMA model is a combination of the AR (Autoregressive) and MA (Moving Average) models as follows:

The autoregressive (AR) model

$$\Psi_t = B_0 + B_1 \Psi_{t-1} + B_2 \Psi_{t-2} + \dots + B_\pi \Psi_{t-\pi} + v_t$$

The moving average (MA) model

$$\Psi_t = X_0 + X_1 v_t + X_2 v_{t-1} + \dots + X_\theta v_{t-\theta}$$

2.3 Data Validation

For the data to be tested using time series models, it has to be stationary. If the raw data is found to be non stationary, it has to go through some transformation first.

2.3.1 Testing stationarity of the time series data

To model the series we check the structure of the data in order to obtain some preliminary knowledge about the stationarity of the series; whether there exist a trend or a seasonal pattern. A time series is said to be a stationary if both the mean and the variance are constant over time. A time plot of the data is suggested to determine whether any differencing is needed before performing formal tests. If the data is non-stationary, we do a logarithm transformation or take the first (or higher) order difference of the data series which may lead to a stationary time series. This process will be repeated until the data exhibit no apparent deviations from stationarity. The times of differencing of the data is indicated by the parameter d in the ARIMA (p,d,q) model. Then an Augmented Dickey-Fuller Test (ADF Test) is used to determine the stationarity of the data.

2.4 Estimation and Order Selection

Box Jenkins method was implemented by observing the autocorrelation of the time series. Therefore, ACF and PACF are core in providing ways to identify ARIMA model. To determine the possible order of the ARMA model, the researchers observed the structure of the



correlogram of the stationary series. To determine which correlations were statistically significant, 95% confidence interval was used. In fitting ARIMA model, the idea of parsimony is important in which the model should have the least number of parameters as much as possible yet still be capable of explaining the series. The more the parameters, the greater noise that can be introduced into the model and hence, the greater standard deviation. In addition, the following methods Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are also applied.

2.5 Model diagnostic checking

Diagnostic checking of the estimated model enable us determine if the estimated model is acceptable and statistically significant. That means, the residuals are not auto correlated and follow normal distribution. The Q statistic of Ljung-Box (1978) was used to test for auto correlation whereas normality test was done using Jarque-Bera (JB) test (1980).

2.6 Forecasting

Forecasting performance of the ARIMA model was determined by computing statistics like; Root mean Squared error, mean absolute error and Theil Inequality Coefficient. The smaller the statistics, the better the model.

III. EMPIRICAL RESULTS

3.1 Test of stationarity

The series under observation was tested for stationarity using ADF test. It was first transformed to logs (abbreviated as LIMPORTS) then tested for stationarity. The null hypothesis was accepted since the series was found to be non-stationary. This conclusion was made based on the P value of the t-statistic (1.000) which was higher than the (5%) level of significance. Findings are presented in Table 1.

Table 1: Stationarity test for logs of time series

Null Hypothesis: IMPORTS has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	3.603527	1.0000
Test critical values:		
1% level	-3.555023	
5% level	-2.915522	
10% level	-2.595565	

*MacKinnon (1996) one-sided p-values.



The series was further transformed to first differences (abbreviated as (DImports), then tested for stationarity. The null hypothesis was rejected since findings showed that the series had attained stationarity as per the P value of the t-statistic (0.000) which was less than the level of significance set at (5%). These findings further illustrated that the series was integrated of order one (abbreviated as I(1). Findings are presented in Table 2.

Table 2: Stationarity test for the first differences of the logs of time series

Null Hypothesis: DLIMPORTS has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.890417	0.0000
Test critical values:		
1% level	-3.557472	
5% level	-2.916566	
10% level	-2.596116	

*MacKinnon (1996) one-sided p-values.

3.2 Order of ARMA Model

To determine the possible order of the ARMA model, the researchers observed the structure of the correlogram of the stationary series. To determine which correlations were statistically significant, 95% confidence interval was used. The 95% confidence interval for the true correlation coefficients for this sample was about +/- 1.96(0.1400) = (-0.264 to 0.264). Correlation coefficients lying outside these bounds are statistically significant at the 5% level. On this basis, ACF and PACF correlations at lags 5 and 13 seem to be statistically significant.

Figure 1: Correlogram of the stationary series

Date: 12/21/16 Time: 12:01

Sample: 1960 2015

Included observations: 55

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
. .	. .	1	0.054	0.054	0.1709
.* .	.* .	2	-0.084	-0.087	0.5882
. .	. .	3	-0.001	0.008	0.5883
. .	. .	4	0.049	0.041	0.7329
** .	** .	5	-0.265	-0.273	5.1333
. *.	. **	6	0.188	0.252	7.3989
. .	.* .	7	-0.045	-0.157	7.5307
. .	. .	8	-0.004	0.070	7.5318
.* .	.* .	9	-0.103	-0.117	8.2479



. *.	.	. *.	.	10	0.160	0.118	10.033	0.438
. .	.	. *.	.	11	0.004	0.083	10.034	0.527
. * .	.	** .	.	12	-0.078	-0.207	10.476	0.574
** .	.	. * .	.	13	-0.284	-0.195	16.508	0.223
. * .	.	** .	.	14	-0.108	-0.206	17.406	0.235
. * 	15	-0.099	0.010	18.171	0.254
. *.	16	0.096	0.055	18.915	0.273
. 	17	0.038	-0.029	19.033	0.327
. 	18	0.028	-0.048	19.099	0.386
. * .	.	. * .	.	19	-0.170	-0.184	21.610	0.304
. 	20	0.024	0.059	21.664	0.359

3.2.1 ARMA (5, 13) (5, 13) Model of DLIMPORTS

Table 3: ARMA (5, 13) (5, 13) Model of DLIMPORTS

Dependent Variable: DLIMPORTS

Method: Least Squares

Date: 12/21/16 Time: 12:06

Sample (adjusted): 1974 2015

Included observations: 42 after adjustments

Failure to improve SSR after 16 iterations

MA Backcast: 1961 1973

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.077814	0.020128	3.865979	0.0004
AR(5)	-0.283557	0.174874	-1.621495	0.1134
AR(13)	-0.419580	0.134454	-3.120613	0.0035
MA(5)	0.338084	0.190828	1.771673	0.0847
MA(13)	0.661832	0.144965	4.565447	0.0001
R-squared	0.510459	Mean dependent var		0.077620
Adjusted R-squared	0.457536	S.D. dependent var		0.160697
S.E. of regression	0.118357	Akaike info criterion		-1.318885
Sum squared resid	0.518307	Schwarz criterion		-1.112020
Log likelihood	32.69659	Hannan-Quinn criter.		-1.243061
F-statistic	9.645263	Durbin-Watson stat		1.640174
Prob(F-statistic)	0.000019			
Inverted AR Roots	.89-.25i .33-.85i .51-.79i -.97	.89+.25i .33+.85i -.51+.79i -.82+.41i	.73+.63i -.14-.94i -.82-.41i	.73-.63i -.14+.94i -.82-.41i
Inverted MA Roots	.92-.25i .34-.88i .53+.82i -1.00	.92+.25i .34+.88i -.53-.82i	.76+.65i -.14-.97i -.85+.42i	.76-.65i -.14+.97i -.85-.42i

Since AR(5) and MA(5) coefficients were not significant, we dropped these from consideration and re-estimated the model with only AR(13) and MA(13) terms, which gave the results in Table 4.



Table 4: ARMA (13, 13) Model of DLIMPORTS

Dependent Variable: DLIMPORTS

Method: Least Squares

Date: 12/21/16 Time: 12:07

Sample (adjusted): 1974 2015

Included observations: 42 after adjustments

Convergence achieved after 15 iterations

MA Backcast: 1961 1973

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.077561	0.020149	3.849319	0.0004
AR(13)	-0.315320	0.093605	-3.368628	0.0017
MA(13)	0.921155	0.026659	34.55313	0.0000
R-squared	0.686395	Mean dependent var		0.077620
Adjusted R-squared	0.670313	S.D. dependent var		0.160697
S.E. of regression	0.092270	Akaike info criterion		-1.859456
Sum squared resid	0.332033	Schwarz criterion		-1.735337
Log likelihood	42.04858	Hannan-Quinn criter.		-1.813962
F-statistic	42.68011	Durbin-Watson stat		1.547970
Prob(F-statistic)	0.000000			
Inverted AR Roots	.89+.22i .32-.86i .52-.75i -.92	.89-.22i .32+.86i -.52+.75i	.68+.61i -.11-.91i -.81+.43i	.68-.61i -.11+.91i -.81-.43i
Inverted MA Roots	.96-.24i .35+.93i -.56-.82i -.99	.96+.24i .35-.93i -.56+.82i	.74-.66i -.12-.99i -.88+.46i	.74+.66i -.12+.99i -.88-.46i

Table 4, shows the coefficient estimates of various autoregressive and moving averages of the Kenyan demand for imports. The coefficients of AR(13) and MA(13), are statistically significant at 5% level of significance.

Based on estimation of these results, the model can be written as follows:

$$DLIMPORTS_t = 0.077561 - 0.315320 DLIMPORTS_{t-1} + 0.921155\epsilon_{t-1} + \epsilon_t$$

t-stat. (3.849319) (-3.368628) (34.55313)

prob. [0.0004] [0.0017] [0.0000]

s.e {0.020149}{0.093605} {0.026659}

Where DLIMPORTSt is import demand, DLIMPORTSt-1 represents import demand at period t-1, ϵ_{t-1} represents the random shock at period t-1, and t is the time period. The estimate for DLIMPORTSt-1 in the equation is negative indicating that the lags are negatively related



to previous variables in the previous period. On the other hand, the estimate for e_{t-1} in the equation is positive indicating that the lags are positively related to previous variables in the previous period.

3.3 Diagnostic Checking of the Model

Diagnostic checking of the estimated model enable us determine if the estimated model is acceptable and statistically significant. That means, the residuals are not auto correlated and follow normal distribution. The Q statistic of Ljung-Box (1978) was used to test for auto correlation whereas normality test was done using Jarque-Bera (JB) test (1980). Findings were presented in the figures below.

Figure 2: Correlogram of the ARIMA (13, 1, 13)

Date: 12/23/16 Time: 10:16

Sample: 1974 2015

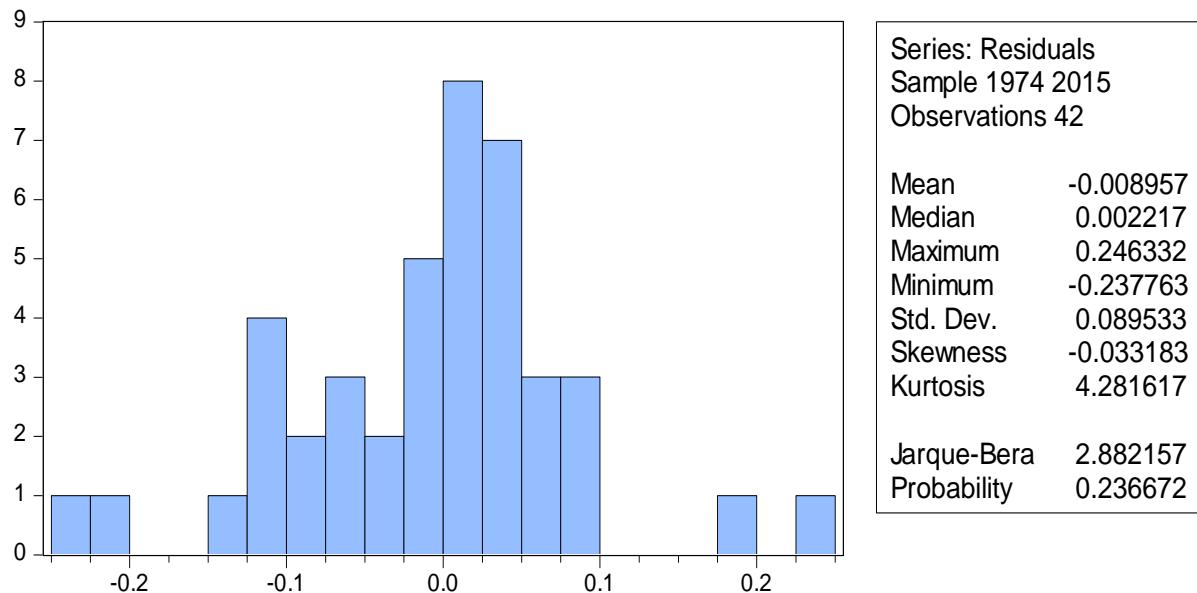
Included observations: 42

Q-statistic

probabilities adjusted
for 2 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
. **	. **	1	0.214	0.214	2.0697
.* .	.* .	2	-0.104	-0.157	2.5661
. *.	. *.	3	0.095	0.166	2.9974
.* .	** .	4	-0.132	-0.237	3.8511
. .	. *.	5	-0.028	0.130	3.8916
. **	. *.	6	0.241	0.158	6.8660
. .	.* .	7	-0.021	-0.097	6.8891
.* .	. .	8	-0.090	-0.021	7.3341
.* .	** .	9	-0.152	-0.241	8.6276
. .	. **	10	0.016	0.264	8.6428
. **	. *.	11	0.241	0.130	12.100
. .	. .	12	0.068	-0.053	12.384
.* .	.* .	13	-0.067	-0.094	12.673
.* .	.* .	14	-0.102	-0.143	13.366
** .	.* .	15	-0.276	-0.077	18.569
.* .	.* .	16	-0.122	-0.086	19.629
. *.	. *.	17	0.186	0.151	22.199
. .	.* .	18	0.011	-0.112	22.208
.* .	.* .	19	-0.149	-0.078	23.999
. .	. *.	20	0.050	0.185	24.211

Figure 3: Distribution of the Residuals of the ARIMA (13, 1, 13) Model



The results of figure 3 indicate that the residuals of ARIMA(13,1,13) model follow normal distribution. Moreover, the results of figure 2 indicate that the Q statistic of Ljung–Box for all the 20 lags has values greater than 0.05 thus the null hypothesis cannot be rejected i.e. there is no autocorrelation for the examined residuals of the series

3.4 Forecasting

The results in figure 6 indicate that the inequality coefficient of Theil has a fairly good forecasting ability. Table 5 below summarizes the forecasting results of the demand for imports over the period 2017 to 2025.

Figure 4: Forecast Accuracy Test on the Model ARIMA (13,1,13)

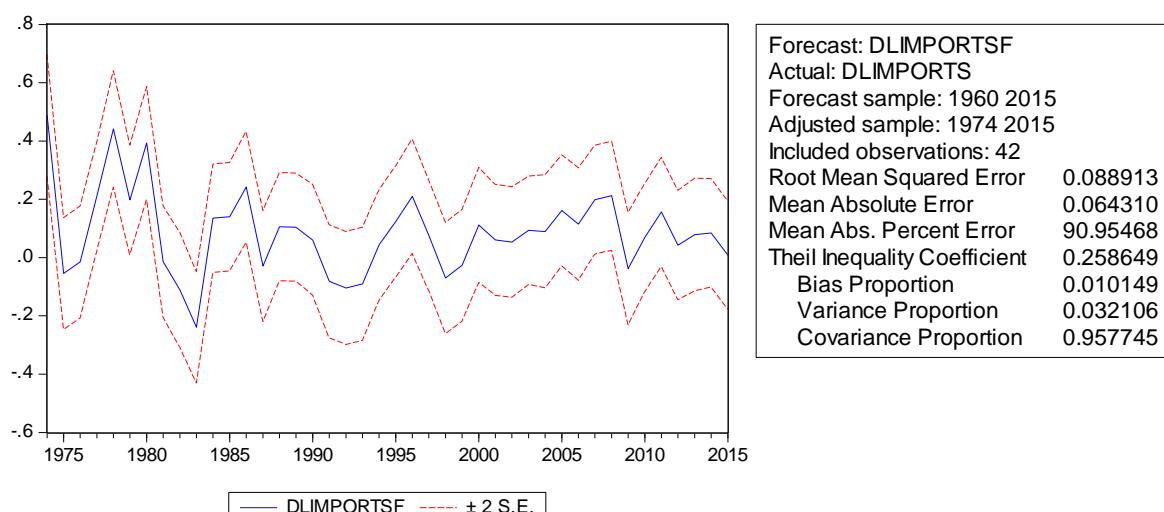




Table 5: The real imports demand forecasts

Year	Value of Imports in Ksh
2017	22824122214.27
2018	25217903035.57
2019	28228120249.88
2020	29903409843.54
2021	31004965054.38
2022	32919626540.22
2023	36007744973.85
2024	39175791249.80
2025	43791428961.98

IV. CONCLUSION AND RECOMMENDATIONS

Using the proposed model, ARIMA (13, 1, 13), the results of forecasting showed that the demand for imports in Kenya have fairly increasing trend over the nine years forecasted. The government of Kenya should evaluate its import composition. Kenya should work on importing raw materials and unfinished goods so as to assemble them locally to cut costs.

REFERENCES

1. MacKinnon, J. G., A. A. Haug, and L. Michelis (1964). "Numerical distribution functions of likelihood ratio tests for cointegration," *Journal of Applied Econometrics*, 14, 563-577.
2. Mwase N. (1990). Foreign Exchange management and allocation systems in Developing countries lesson of African Experience, *African Development Review*, pp 65-78.
3. Robert.C (2005). *International Economics*. Ohio: Thomson Corporation.
4. Osoro, K. (2012). Kenya's Foreign Trade Balance: An Empirical Investigation, *European Scientific Journal*, vo1.9 pp 19.No.6, United Nations Development Programme
5. Republic of Kenya (2007). *The Kenya Vision 2030*. Nairobi: Government Printers.