

## Behaviour of Construction Output in Kenya: *An ARIMA Model*

Emmanuel Thyaka Mbusi\*, Titus Peter Kivaa and Willy Mwangi Muturi

Received on 28th August, 2020; Received in revised form 2nd November, 2020; Accepted on 16th November, 2020.

### Abstract

*In a recent investigation of ways to enhance the role of construction industry in the economy of Kenya, methods of modeling the construction activity were explored. This is because there remains a problem in the construction industry failing to meet demand for constructed facilities in the country at the macro-level. In the exploration, it was noted that ARIMA (Autoregressive Integrated Moving Average) models have widely been used in modeling construction industry – either as a whole or the different sectors of the industry - around the world. This modeling approach which is also popularly known as the Box-Jenkins approach was adopted in this study. This paper is a report of the study. Construction output data collected and analyzed in this paper was obtained from the Kenya National Bureau of Statistics (KNBS) for the period between 1977 and 2019 – 43 years. The developed ARIMA model was evaluated on its predictive power on the basis of an out-of-sample forecast. The study adopted two measures of accuracy - MAPE and RMSE - which produced fairly good results. Therefore, the model can be used for forecasting construction output in Kenya. Finally, it is recommended that the model be applied in setting construction production targets at the macro-level and guiding policy formulation to stimulate higher performance of the construction industry in Kenya.*

**Keywords:** Box-Jenkins approach, Construction industry, Construction output, Forecasting.

### INTRODUCTION

In Kenya, there is a number of policy documents designed to provide guidance on aspired targets in the provision of adequate infrastructure in the country. Examples are the National Housing Policy, the Big Four Agenda and Kenya Vision 2030. However, it has been observed that either the production targets implied in these policy documents have been too high or the existing policy meant to propel the industry towards achievement of the said targets is not proper. The research sought to find out exactly what influences construction output levels in Kenya. There is the need to develop models for explanation and prediction of construction output in Kenya, to inform policy design, implementation and evaluation.

A realistic start point is the ARIMA (Autoregressive Integrated Moving Average) model. This is because construction output is a time series variable, and can therefore be modeled as a self-projecting

variable. In this case, the level of the variable at a given point in time is explained by its level(s) in the previous time(s). That is the concept behind the ARIMA model.

The ARIMA method of time-series analysis was initially developed by Box and Jenkins in 1976 and has become popularly known as the Box-Jenkins approach (Hua & Pin, 2000). In time series analysis, ARIMA modeling is an alternative to multiple regression modeling - whereby the level of a time series variable at a given point in time is explained by levels of other (or independent) variables.

In this study, the research objective was to formulate an ARIMA model of construction output in Kenya and eventually use it to forecast the output. It was postulated that the level of annual construction output in a given year was influenced by the past years and stochastic error terms.

\*Corresponding author:

Emmanuel Thyaka Mbusi, School of Engineering and Built Environment, Kirinyaga University, Kenya.

Email: [thyakambusi@gmail.com](mailto:thyakambusi@gmail.com)

## THEORY

Construction industry plays numerous roles in the economy of a country. In Kenya particularly, the industry is a key contributor to the Gross Fixed Capital Formation (GFCF), which is a statistical measure of the value of all the acquisitions of fixed assets, whether fresh or existing. On average, the construction industry contributes 40% of the GFCF, and 4% of the Gross Domestic Product (GDP) in Kenya. Also, it is estimated that people in the region of 80,000 are employed by the construction industry at any one moment (Mitullah & Wachira, 2003). All the same, there are notable fluctuations in the level of percentage contribution of construction activity on the GFCF. **Table 1** gives a snapshot of that variability, over the period 2000 to 2004.

**TABLE 1:** Percentage contribution to GFCF of Construction Industry in Kenya

Year	Contribution to total GFCF (%)
2000	47.9%
2001	46.4%
2002	47.8%
2003	53.0%
2004	51.6%

Source: K'Akumu 2007

The construction industry's percentage contribution towards the GFCF rose constantly from year 2000, although it went down a bit in the year 2001. It picked up the rise again in the period 2002 to 2004. This indicates that there were well-sustained construction activities in the country during that period. However, it leaves out a number of things unexplained. For example, it does not indicate the extent to which construction activity level - or percentage contribution - in a given year was influenced by its levels in the previous year(s). This calls for the more intense analysis of the construction GFCF variable.

### Application of ARIMA models in the study of Construction Industry

ARIMA models are widely used in the modeling and forecasting of activity levels in different

sectors of construction industry. Examples of ARIMA usage in the forecasting of housing, offices and factories are found in the following research reports: Akintoye & Skitmore (1994); Bickerton & Grunerberg (2013); Hua & Pin (2000); Kivaa (2008); and Notman, Norman, Flanagan & Agapiou (1998). While Akintoye & Skitmore (1994) and Bickerton & Grunerberg (2013) are examples of researchers who have studied construction industry in the UK, Hua & Pin (2000), is an example of a researcher who has studied construction industry in Singapore. For Kenya, ARIMA modeling was used in Kivaa (2008), in studying construction output in the country from 1963 – 2003. However, the predictive power of this model was rather low.

The diversity of the above-listed examples is a pointer to the fact that there is growing interest towards application of ARIMA modeling in the study of Construction Industry. Apparently, continual adoption of ARIMA modeling in the study of construction activity in Kenya should eventually produce a concrete set of decision tools to aid property developers and policy makers in the construction industry of Kenya. It is for that reason that ARIMA methodology was adopted in this study.

## RESEARCH METHODS

In brief, an ARIMA model is a univariate or a single vector model. As highlighted before, it is a technique mostly used for projecting future values of a time series on the basis of its own inertia. It is observed to perform much better than alternative forecasting methods, if applied on short-term forecasts. Creation of a credible ARIMA model requires at least forty (40) data points of historical data. According to Morrison (2020), for example, researchers are better off adopting other forecasting methods if their data points are below thirty-eight (38) in number. In this study, the data points were 43 – i.e. 1977 to 2019. For that reason, the data collected and analyzed in the study was adequate for ARIMA modeling of the construction output in Kenya.

The ARIMA regression model for construction output (CO) in Kenya may be statistically expressed using the following model: -

$$CO_t = \theta + \alpha_1 CO_{t-1} + \alpha_2 CO_{t-2} + \dots + \alpha_p CO_{t-p} + \beta_0 \mu_t + \beta_1 \mu_{t-1} + \beta_2 \mu_{t-2} + \dots + \beta_q \mu_{t-q} \dots \dots (1)$$

Where:

$\theta$  = represents a constant term

$p$  = the number of autoregressive terms

$q$  = the number of moving average terms

$\mu_t$  = unautocorrelated random error term with zero mean and constant variance ( $\sigma^2$ ).

In this model, the level of construction output at time  $t$  ( $CO_t$ ) depends on its level in the previous  $p$  time periods (years), and on the moving average of the current and past  $q$  error terms.

### Evaluation of Model Accuracy

The accuracy of the ARIMA model was evaluated using the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE). The first forty observations were used as the modeling data, whereas the last three were used for testing the model for forecasting accuracy.

## RESULTS

### The Time-Series Data

Data was obtained from the Kenya National Bureau of Statistics (KNBS) for the period 1977

to 2019 – i.e. 43 years. The time series data was first checked for stationarity and found to be non-stationary, just like most other economic data. Consequently, the data was transformed in a bid to achieve stationarity and avoid spurious or nonsensical regression results. According to Gujarati and Porter (2009), it is generally assumed that all time series data used for empirical work are stationary. In order to achieve stationarity, the first differences of the construction output data were computed and tested for stationarity. They were found to be stationary and subsequently used for the analysis.

### Stationarity test for the data

A plot of the first differences of the logarithm of construction output was done to give the pictorial view of the transformed construction output data. The results were as shown in **Figure 1**. Indeed, the first differences produce a stationary time series. In order to be more confident in the deductions, further stationarity tests were carried out applying the Augmented Dickey-Fuller (ADF) unit root test and the results are displayed on **Table 2**.

The ADF unit root test method gave good results as observed in **Table 2**. The null hypothesis that the variable had a unit root was rejected at 95% confidence interval. This assured the researcher that further analysis can be carried out using the first differences of this variable.

**TABLE 2:** Unit Root Test for First Differences of Construction Output

Null Hypothesis: D(CO) has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 4 (Automatic - based on SIC, max. lag=9)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-13.16401	0.0000
Test critical values:	1% level		-4.252879	
	5% level		-3.548490	
	10% level		-3.207094	
*MacKinnon (1996) one-sided p-values.				

Source: Authors 2020

**Model Identification and Estimation**

An ARIMA model identification process was carried out by first assessing the autocorrelation and partial correlation of the variable's data using a correlogram of the first differences of logarithm of construction output. **Table 3** displays the results.

From **Table 3**, it did not appear as if there was any systematic information in the construction output time series data that could help in explaining its behavior. The autocorrelation as observed from this correlogram lies within its boundaries, which means that even the fourth and the fifth order autocorrelations were not significant. For this reason, it is clear that it does not have any theoretical resemblance of an autoregressive (AR) process nor a moving average (MA) process.

It was further drawn from model selection criteria, Akaike Information Criteria (AIC), as shown in **Table 4** and **Figure 2**, that numerous regressions were run of different combinations of AR and MA processes, and tests of their explanatory powers carried out. This position is also reaffirmed by forecast comparison graph which appears in **Figure 3**.

In the final analysis, ARIMA (0, 1, 0) model which yielded some relevant results was modeled. It is therefore this ARIMA (0, 1, 0,) process which gives a comprehensive description of the annual variations of construction output in Kenya between 1977 and 2019. This carries the implication that ARIMA (0, 1, 0) process has no AR and MA terms and that construction output data levels are integrated of order one. These regression results of the ARIMA (0, 1, 0) are presented on **Table 5**.

The R<sup>2</sup> value observed for this ARIMA model is zero though it is quite important to note that this is due to the absence of AR and MA values in the model. As observed from the regression output in **Table 5**, it is as well observed that the Durbin Watson (DW) value indicates that residuals do not have serial correlations. Further, it is however notable that the constant term is insignificant which implies that all construction output variations in Kenya were well explained in this model as per the analyzed data of 1977 to 2019. This state of affairs was demystified by the correlogram of the residuals as shown in **Table 6**. From **Table 6**, it was observable that all the autocorrelations in different lag-lengths were all kept within the required limit.

TABLE 3: Correlogram of Differenced Logarithm of Construction Output

Date: 07/29/20 Time: 08:53  
 Sample: 1977 2016  
 Included observations: 39

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.094	-0.094	0.3706	0.543
		2 0.026	0.018	0.4008	0.818
		3 -0.054	-0.050	0.5296	0.912
		4 0.299	0.292	4.6026	0.331
		5 -0.328	-0.306	9.6505	0.086
		6 0.173	0.170	11.097	0.085
		7 -0.009	0.012	11.101	0.134
		8 -0.024	-0.155	11.132	0.194
		9 -0.152	0.064	12.357	0.194
		10 0.238	0.058	15.489	0.115
		11 -0.165	-0.120	17.042	0.107
		12 -0.120	-0.098	17.898	0.119
		13 -0.117	-0.151	18.739	0.131
		14 0.019	-0.107	18.763	0.174
		15 -0.106	0.086	19.505	0.192
		16 0.035	-0.052	19.592	0.239

Source: Authors 2020

TABLE 4: Model Selection Criteria table results

Dependent Variable: DLOG(CO)				
Date: 08/04/20 Time: 15:14				
Sample: 1977 2016				
Included observations: 39				
Model	LogL	AIC*	BIC	HQ
(0,0)(0,0)	-27.079340	1.491248	1.576559	1.521857
(1,1)(0,0)	-25.774731	1.526909	1.697531	1.588127
(1,0)(0,0)	-26.910710	1.533883	1.661849	1.579796
(0,1)(0,0)	-26.916896	1.534200	1.662166	1.580113
(4,2)(0,0)	-22.188398	1.548123	1.889366	1.670558
(3,2)(0,0)	-23.335642	1.555674	1.854262	1.662805
(0,4)(0,0)	-24.485667	1.563368	1.819300	1.655194
(1,3)(0,0)	-24.520258	1.565141	1.821074	1.656968
(4,1)(0,0)	-23.618845	1.570197	1.868785	1.677328
(1,4)(0,0)	-23.624602	1.570492	1.869080	1.677623
(1,2)(0,0)	-25.656015	1.572103	1.785380	1.648625
(2,1)(0,0)	-25.707113	1.574724	1.788001	1.651246
(2,2)(0,0)	-24.848803	1.581990	1.837922	1.673816
(3,4)(0,0)	-21.903387	1.584789	1.968688	1.722529
(2,0)(0,0)	-26.904921	1.584868	1.755489	1.646085
(0,2)(0,0)	-26.909873	1.585122	1.755743	1.646339
(2,3)(0,0)	-23.984481	1.588948	1.887536	1.696079
(3,1)(0,0)	-25.133612	1.596596	1.852528	1.688422
(2,4)(0,0)	-23.209481	1.600486	1.941730	1.722921
(4,0)(0,0)	-25.268700	1.603523	1.859456	1.695349
(3,0)(0,0)	-26.858699	1.633779	1.847057	1.710301
(0,3)(0,0)	-26.909847	1.636402	1.849680	1.712924
(4,4)(0,0)	-22.000629	1.641058	2.067612	1.794102
(4,3)(0,0)	-23.050255	1.643603	2.027502	1.781342
(3,3)(0,0)	-24.140351	1.648223	1.989467	1.770658

Source: Authors 2020

Based on regression results displayed in Table 5, the following is an ARIMA expression that describes Kenya's construction output: -

$$DLog(CO_t) = 0.21 \dots\dots\dots (2)$$

Where:

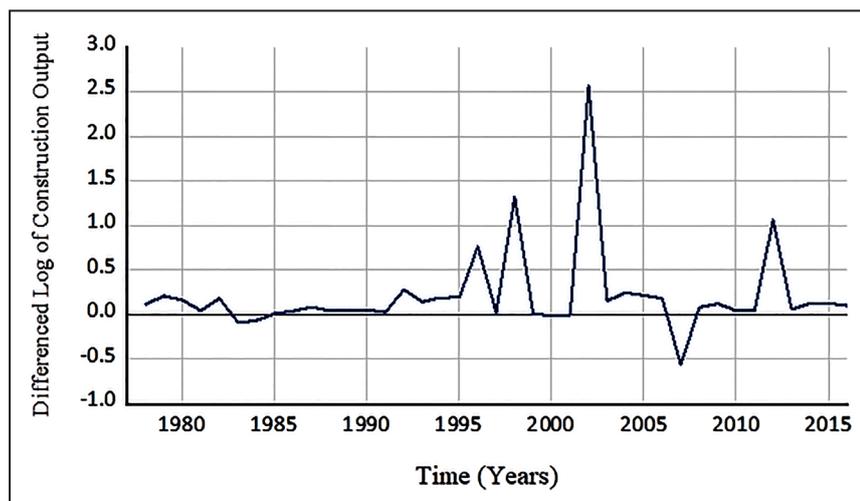
$$Dlog(CO_t) = Log(CO_t) - Log(CO_{t-1}) \text{ (the first difference of logarithm of construction output)}$$

The ARIMA equation can now be expressed in terms of construction output level and given as hereafter: -

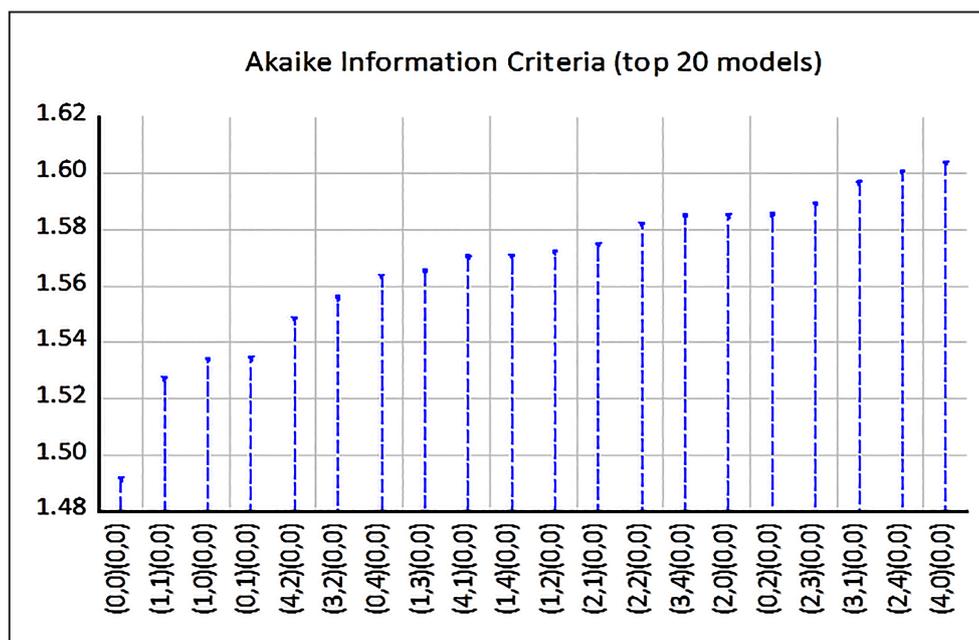
$$DLog(CO_t) = 0.21 \text{ is now expanded to } Log(CO_t) - Log(CO_{t-1}) = 0.2$$

Reorganization of the equation above results in the following equation: -

$$Log(CO_t) = 0.21 + Log(CO_{t-1}) \dots\dots\dots (3)$$



**FIGURE 1**  
Differenced Natural Logarithm of Construction Output  
Source: Authors 2020



**FIGURE 2**  
Akaike Information Criteria results  
Source: Authors 2020

This equation implies that Kenya’s construction output at any one given year is influenced by its level in the previous year.

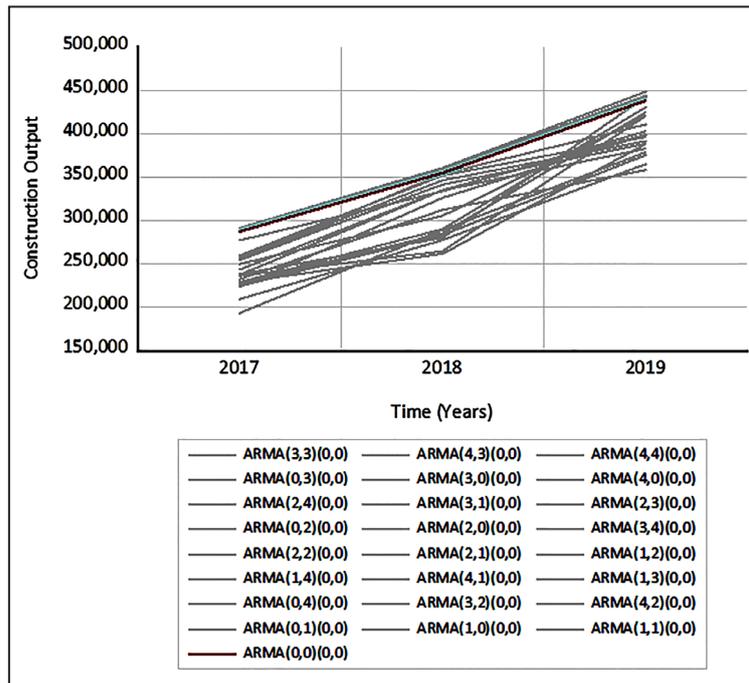
**ARIMA Forecasting**

The forecasting was carried out to predict the levels of construction output in the next three

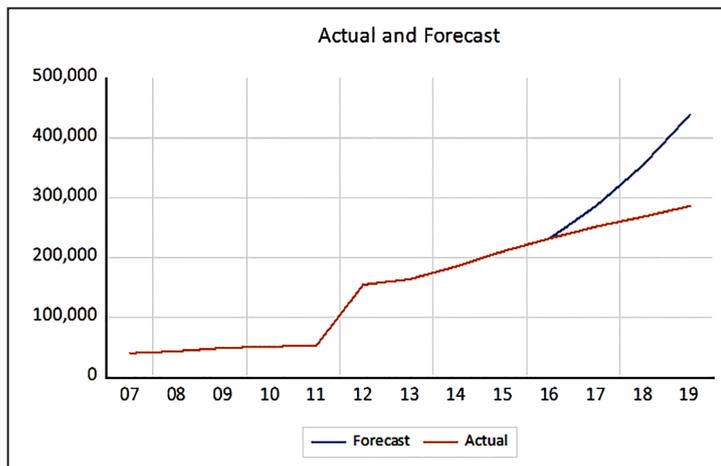
years. That is from 2017 to 2019. The forecasting results are shown in **Table 7** and **Figure 4**.

**Evaluation of Forecasting Accuracy**

Forecasting accuracy for the three (3) year out of sample forecasts were evaluated and the results obtained are given in **Table 8** and **Figure 5**. The



**FIGURE 3**  
Forecast comparison graph  
Source: Authors 2020



**FIGURE 4**  
ARIMA forecasting graph  
Source: Authors 2020

results appear impressive since the accuracy measurement parameters show appreciable figures. The figures as presented in Table 8 and forecast comparison graph show that the RMSE is 103098.5 and the MAPE is 23.8%. These error levels imply that the ARIMA model developed

in this research carries a fairly good predictive power and has explained the construction output variations in Kenya for the period in question - 1977 to 2019 - to a fairly good degree of accuracy. The model is realistic.

**TABLE 5:** ARIMA Model of Differenced Logarithm of Construction Output

\*Dependent Variable: Difference of Logarithm of Construction output

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.211950	0.078600	2.696582	0.0104
R-squared	0.000000	Mean dependent var		0.211950
Adjusted R-squared	0.000000	S.D. dependent va		0.490854
S.E. of regression	0.490854	Akaike info criterion		1.439966
Sum squared resid	9.155627	Schwarz criterion		1.482622
og likelihood	-27.07934	Hannan-Quinn criter.		1.455271
Durbin-Watson stat	2.185191			

Source: Authors 2020

**TABLE 6:** Correlogram of Residuals

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.087	-0.087	0.3441	0.557
		2 0.031	0.024	0.3895	0.823
		3 -0.049	-0.045	0.5029	0.918
		4 0.302	0.297	4.9430	0.293
		5 -0.333	-0.315	10.486	0.063
		6 0.164	0.163	11.867	0.065
		7 -0.016	0.002	11.880	0.105
		8 -0.017	-0.148	11.895	0.156
		9 -0.144	0.081	13.049	0.160
		10 0.253	0.074	16.743	0.080
		11 -0.147	-0.097	18.032	0.081
		12 -0.104	-0.085	18.704	0.096
		13 -0.113	-0.145	19.520	0.108
		14 0.023	-0.103	19.554	0.145
		15 -0.135	0.051	20.795	0.144
		16 0.005	-0.083	20.796	0.186
		17 -0.082	-0.086	21.297	0.213
		18 -0.062	-0.052	21.596	0.250
		19 -0.065	-0.075	21.936	0.287
		20 -0.001	-0.073	21.937	0.344

Source: Authors 2020

**TABLE 7:** Automatic ARIMA forecasting results

Automatic ARIMA Forecasting
Selected dependent variable: DLOG(CO)
Date: 08/04/20 Time: 15:14
Sample: 1977 2016
Included observations: 39
Forecast length: 3
Number of estimated ARMA models: 25
Number of non-converged estimations: 0
Selected ARMA model: (0,0)(0,0)
AIC value: 1.49124819017

Source: Authors 2020

## DISCUSSION

ARIMA modeling of construction output in Kenya is not a common occurrence. From the literature reviewed it was applied in Kivaa (2008), who modeled the construction output from 1963 to 2003. The findings in that research work were not encouraging; the MAPE was 40%, which was much higher than the MAPE of 23.80% realized in this study. This means that the ARIMA model developed in this study is a refinement of the ARIMA model previously developed. Therefore, ARIMA modeling can now be used with more confidence in explaining and forecasting construction output in Kenya. These findings are a new dawn for the formulation or evaluation of construction industry policy in Kenya.

Moreover, the prediction power of the model developed in this study is fairly high. The MAPE is 23.8% and compares well with prediction powers of similar models developed elsewhere, for example in Hua & Pin (2000). This model is therefore realistic and usable. All the same, further study should be done in order to enhance the predictive power and compare it with prediction powers of alternative prediction models.

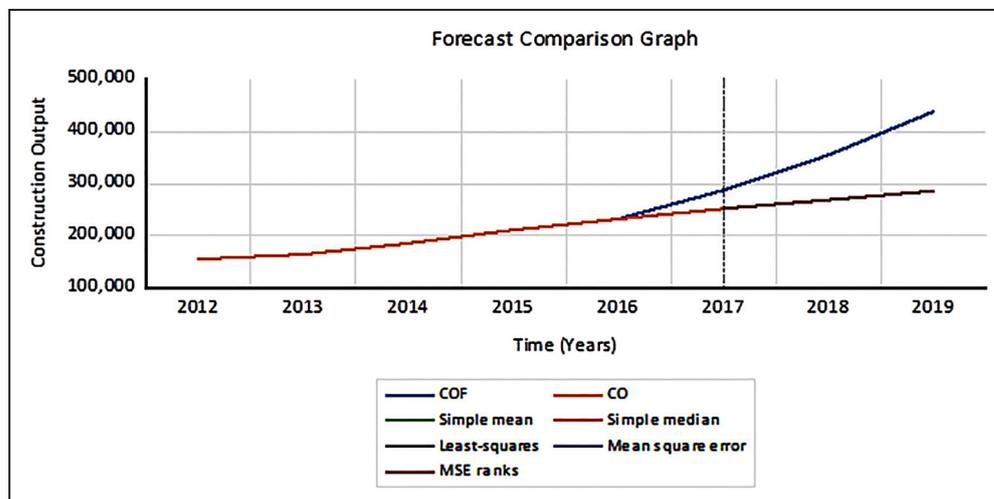
## CONCLUSION AND RECOMMENDATIONS

ARIMA modeling for construction activity in Kenya is a realistic endeavour. Using ARIMA modeling, the researchers in this study developed forecasting model for construction output in Kenya. The model can be expressed as follows: -

**TABLE 8:** Forecasting evaluation results

Forecast Evaluation						
Sample: 2017-2019						
Included observations: 3						
Evaluation sample: 2017-2019						
Training sample: 2000-2016						
Number of forecasts: 6						
Combination tests						
Null hypothesis: Forecast includes all information contained in others						
Forecast	F-stat	F-prob				
CO	NA	NA				
Evaluation statistics						
Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
CO	103098.5	91235.61	23.75538	27.57530	0.162413	1.569534
Simple mean	103098.5	91235.61	23.75538	27.57530	0.162413	1.569534
Simple median	103098.5	91235.61	23.75538	27.57530	0.162413	1.569534
Least-squares	103098.5	91235.61	23.75538	27.57530	0.162413	1.569534
MSE ranks	103098.5	91235.61	23.75538	27.57530	0.162413	1.569534

Source: Authors 2020



**FIGURE 5**  
Forecast comparison graph  
Source: Authors 2020

$$\text{Log}(CO_t) = 0.21 + \text{Log}(CO_{t-1})$$

Where:

$\text{Log}(CO_t)$  is the natural logarithm of construction output at time  $t$  in Kenya Shillings.

The prediction power of the model is fairly high; the MAPE is 23.8%. This equation implies that construction output in Kenya at any one given year is most significantly influenced by its level in the one previous year.

In the light of the foregoing, it is recommended that policy makers in Kenya today should increase construction activity in the country for next year(s) by increasing construction activity in the current year ( $CO_t$ ). That calls for the decisive boosting of investment in real estate and infrastructure development coupled with increased organizational effectiveness of the industry. For the country to achieve the targets of the Big Four Agenda and Vision 2030, it cannot be business as usual for the construction industry in Kenya.

## CITED REFERENCES

**Akintoye, A. & Skitmore, M. (1994).** Models of UK Private Sector Quarterly Construction Demand. *Construction Management and Economics*. 3 - 13.

**Bickerton, M. & Grunerberg, S.L. (2013).** The London Interbank Offered Rate (LIBOR) and UK Construction Industry Output 1990 - 2008. *Journal of Financial Management of Property and Construction*. 18, 268 - 281.

**Gujarati, D.N. & Porter, D.C. (2009).** *Basic Econometrics*. New York: McGraw-Hill Education.

**Hua, G.B. & Pin, T.H. (2000).** Forecasting construction industry demand, price and productivity in Singapore: The Boxjenkins Approach. *Construction Management and Economics*. 607 - 618.

**K'Akumu, O. (2007).** Construction Statistics Review for Kenya. *Construction Management and Economics*. 315 - 326.

**Kivaa, T.P. (2008).** *A System Dynamics Model of Construction Output in Kenya*. Melbourne: Unpublished.

**Mitullah, W.V. & Wachira, I.N. (2003).** *Informal Labour in the Construction Industry in Kenya*.

**Morrison, J. (2020).** *ARIMA*. Retrieved May 9, 2020 from Arima: <https://www.forecastingsolutions.com/arima.html>.

**Notman, D., Norman, G., Flanagan, R. & Agapiou, A. (1998).** A time-series analysis of UK annual and quarterly construction output data (1955-95). *Construction Management and Economics*. 409 - 416.