The International Journal Of Engineering And Science (IJES)
||Volume||2 ||Issue|| 9||Pages|| 07-18||2013||
ISSN(e): 2319 – 1813 ISSN(p): 2319 – 1805

Time Series Analysis Model for the Rate of Influx of Refugees in Kenya

[1]Mudanyi Lenard Allan, [2]Mutua Nicholas Muthama, [3]Sinino Zipporah Kinyanjui
[1,2,3]Department of Mathematics and Informatics, School of Science and Informatics, Taita Taveta University College

ABSTRACT

There has been a large influx of refugees in Kenya from the neighboring countries due to their political instabilities. In this study, the number of refugees entering Kenya every year for a period 1993 to 2010 was analyzed using time series methods. Autocorrelation and partial autocorrelation function are calculated for the collected data. The appropriate Autoregressive Integrated Moving Average (ARIMA) model was fitted. The validity of the developed model was then checked using standard statistical techniques. Since the ARIMA model has a forecasting power, it was used in forecasting the number of refugees in Kenya.

Keywords: *Influx, Refugees, Time series, Autocorrelation, Forecasting.*

Date of Submission: 31 July, 2013 Date of Acceptance: 10, September 2013

I. INTRODUCTION

It is sad to note that in 21st century many countries in Africa are still experiencing civil wars. These civil wars have led to lack of quality essential commodities like food, clean water. As a result of this citizen of these countries continue to flee to other countries in search of those essential commodities. An oxford English dictionary defines a refugee as a person who flees his country for safety purposes. Since Kenya is one of politically stable countries in the sub-Saharan Africa, there has been influx of number of refugees from different countries in eastern central Africa. Kenya has seen a large-scale influx of refugees, mostly triggered by the protracted humanitarian crisis in the neighboring countries. The countries from which the refugees come from are Somalia, Ethiopia, Rwanda, Burundi, Congo, Sudan and Eritrea. According to annual report produced by UNHCR in 2009, Kenya is among the top ten countries in the world hosting the largest number of refugees. Due to this, in 2006, the government of Kenya passed a Refugee Act implementing the 1951 United Nations Convention Related to the Status of Refugees, the 1967 Protocol and the 1969 OAU Convention. The development of the Act followed a period of sustained advocacy by UNHCR and civil society organizations, including RCK. In addition to this, there are several nongovernmental organizations which are concerned with helping these refugees who are resettled in refugee camps such as Daadab and Kakuma. These organizations include; African Development and Emergency Organization, CARE International, UNHCR among others.

This study was aimed at establishing the trend by which refugees have been entering Kenya, and forecast the number in some years to come. The data was collected for a certain period of time to establish the trend that the collected data follows. The ARIMA model of time series analysis was developed and was used in forecasting. The aspects of time series that were studied included autocorrelation, trend or seasonal variation among others. In time series, trend is a long term movement in a time series. It is the underlying direction either an upward or downward tendency and rate of change in a time series when allowance has been made for the other components. Using the trend extrapolation can be obtained. Extrapolation is a method of estimating the value of the variable at times which is yet to be observed. Trend can also be removed from the time series by a method called differencing, which leads to provision of clear view of the true underlying behavior of the series. The other aspect of time series that is to be studied is autocorrelation, which is the relationship or correlation between members of a time series observations, and the same values at a fixed time interval. Autocorrelation occurs when residual error terms from observations of the same variable at different times are related. Since data is studied for some period of time, autocorrelation is established among the collected data. Lag values are studied in order to establish autocorrelation.

1.1 STATEMENT OF THE PROBLEM.

Kenya as a country has been experiencing several problems it the recent past. The problems include; human trafficking, insecurity, money laundering, and recently unstable population rise especially in the Northern Eastern part that even led to the cancellation of 2009 census result. It is believed that these problems are related to large number of refugees. This study is aimed at establishing the trend at which refugees have entering this country.

1.2 JUSTIFICATION OF THE STUDY

The purpose of performing this study is to establish the pattern of entry of refugees and asylum seekers in Kenya, after which the ARIMA model of time series is constructed. This model is important in that it helps in forecasting of the number of refugees in some years to come. As a result the government and relevant organizations are advised to plan on how to help countries from which the refugees have been coming from. Since the data collected is assumed to have some internal structure, the statistic distribution which the collected data follows is established. This can be of greater necessity in future.

1.3 OBJECTIVE OF THE STUDY

The objectives for which this study is aimed at accomplishing are;

- To study the statistical distribution generated from the collected data.
- To use time series analysis to project the number of refugees Kenya will be hosting by the year 2030 in order for the relevant authorities to be advised on what measures to be taken for this burden not to derail achievement of vision 2030.

1.4 RESEARCH QUESTIONS AND HYPOTHESIS

Research questions which are answered in this study are the following;

- Does statistical distribution for the collected data normal distribution?
- How is the trend of influx of refugees?

In this study the following two null hypotheses are formulated

Hypothesis 1;

H0: the statistical distribution is normal

Hypothesis 2;

H0: the trend of influx of refugees is upward

1.5 ASSUMPTIONS OF THE STUDY

- In Time-Series Models we assume that we know nothing about the causality effects on the variable we are trying to forecast, but we study the past behavior of the time series in order to predict the future.
- Time series methods were known to take into account that data collected for some period of time had some internal structure.

1.6 LIMITATIONS OF THE STUDY

The data collected is of secondary nature hence the data does not have high level of accuracy.

II. LITERATURE REVIEW

2.1 INFLUX OF REFUGEES

There are various studies carried out on influx of refugees across the world by different researchers of various institutions. These studies focus on the effects of influx of refugees on various aspects such as economy, environment, and security among others. The study by Idean, Salehyan and Kristian Skrede Gleditsch of International Organization (2006) on refugees was entitled ‘Refugees and the Spread of Civil War’. This research highlighted the contribution of influx of refugees to the civil war experienced in the country where they go. Idean and Kristian argued that regions which receive more refugees have had higher level of civil war. They gave example of regions such as Central America, the Great Lakes region of Africa, and South-East Asia, which have witnessed numerous civil wars within several states, whereas other areas such as Europe and the Southern Cone of Latin America have had a relatively low frequency of internal conflict because there a lower rate of influx of refugees in those areas.

The research by Sesay, Fatmata and Lovetta (2004) highlighted the effects of influx of refugees on the economy. In this study, it was argued that the flow of refugees and the prevalence conflict in developing countries was introduced as key variables among others as possible factors that could explain the rate of economic growth in Africa and other regions of the developing world. He said that the poorest countries have provided asylum and shelter for almost three quarter of the world's refugees over the past decade. However he found out that statistics on refugees in these countries suffer from unavailability, periodical lags and sometimes contradictions. The study also used standard methods of panel data estimation (fixed effects and random effects) which made it possible to control for time-invariant country-specific effects in further establishing short run effects of refugees and conflict in developing countries. Empirical analysis was also carried out, with the use of pooled cross sectional data, on factors affecting the flow of refugees, the effects of refugees on developed and developing countries, and factors attracting/distracting refugees to /from certain developing countries. The study also looked at the spillover effects of conflicts in neighboring countries.

A request by the government of Guinea allowed UNEP in close cooperation with UNHCR to carry out study on the impacts of refugees on the environment (1999). This followed after refugees from Liberia and Sierra Leone had been forced to flee their respective countries due to civil war. In that research it proved to be very difficult to accurately estimate the total number of refugees, as not all refugees were registered, and many were living outside the UNHCR refugee's camps. However, it was estimated that over 300,000 refugees lived in the rural areas in Southern Guinea, while a same number were living in the urban centres, mainly in the south and also in Conakry. The study entitled 'Impact of refugees in Northwestern Tanzania' (2003) was conducted by the center for study of forced migration, University of Dar es salaam. The people involved were Dr. Bonaventure Rutinwa, Dr. Khoti Kamanga and Karren L. K. Washoma. In their study they looked at the impacts of in refugees on various aspects such as external and internal security, environment, physical and social infrastructure, local governance and administration, and socio-economic development. The objective of this study was to gather and analyze qualitative and quantitative data regarding the above subject for policy and advocacy guidance. In conclusion they found out that refugees had negative impacts on all aspects mentioned above. 'Hidden an exposed refugees in Nairobi, Kenya' (2010) is a study carried out by Sara Pavanello, Samir Elhawary and Sara Pantuliano of Humanitarian Policy Group. The paper produced by these researchers aimed at finding out the number of refugees in Nairobi. This was done because there had been a number of refugees that had escaped from refugees; camps in Daadab and Kakuma.

2.2 TIME SERIES MODELS

Time series has evolved through various models to the presently used models. Decomposition model is among the oldest models of time series, but this model had theoretical weaknesses from the statistical point of view. This was followed by the crudest form of forecasting called moving averages method. Due to the weaknesses of moving average method, there was need for improvement of this method that had equal weights, so exponential smoothing methods were developed, and gave more weights to recent data. But exponential smoothing methods were just recursive methods without any distributional assumptions about error structure in them. Hence the need for improvement to get rid of the limitations of exponential smoothing, that led to development of particular cases of statistical sound Auto-Regressive Integrated Moving Average (ARIMA) models. In early 1970's Box and Jenkin championed in developing methodologies for time series modeling in the univariate case which was known as Univariate Box-Jenkins (UBJ) ARIMA modeling. From this development, various organizations have managed to come up with various models based on time series. These models included; viz time series decomposition model, seasonal ARIMA model, vector ARIMA models using multivariate time series, ARIMAX models i.e. ARIMA with explanatory variables There have many research and work done using various model developed based on time series. The applications of time series have been done on research works that involved studying data relationships and using the data to forecast the future

Mr. Robert Engle was awarded Nobel Prize of economics for developing a time series model that was used in analysis of finance (2003). His model was called Autoregressive Conditional Heteroscedastic (ARCH) model, and it had been pioneered in 1982. This model was then modified in 1986 by Tim Bollerslev to Generalized ARCH (GARCH) model that is considered today to be the most often applied-varying model. These two models are autoregressive models that are used to measure volatility of assets. The importance of volatility modeling is attributable to a basic observation about asset return data: large returns tend to be followed by more large returns. The applications of volatility are pricing derivatives and risk Management

Willem Dekker of the Institute for Fisheries Research, RIVO in Netherlands used time series to develop a model known as Length-based Virtual Population Analysis (LVPA) (1993), which was used specifically to assess Lake Fisher IJ sselmeer eeies in terms of state of the stock in relation to notional targets such as growth overfishing and recruitment overfishing. This model started as Virtual Population Analysis but it later adapted lengthy structured data. P. R. Crone, R. J. Conser and J. D. McDaniel used time series to develop a model called Age Structured Assessment Program (ASAP) (2004). This model was used in fish population analysis of North Pacific Albacore based on an age structure. This model relied on a Virtual Population Analysis (VPA) model for purposes of developing an international-based consensus regarding the 'status' of this fish stock, which largely serves as the scientific information for guiding potential management in the future. That is, the modeling research presented here was intended to serve as an alternative population analysis based on a forward-simulation, catch-at-age model (ASAP) and ultimately used to evaluate more fully the relationship between this species' population dynamics and associated fishery operations such as areas of uncertainty in an overall stock assessment, than was possible using a more general assessment approach, such as a VPA. The ARIM A model for forecasting oil palm price (2006) in Thailand was developed by Rangsan Nochai and Titida Nochai. Their research was studying the model of forecasting oil palm price of Thailand in three types, i.e. farm price, wholesale price and pure oil price for the period of five years, from 2000 – 2004. The objective of the research was to find an appropriate ARIMA Model for forecasting in three types of oil palm price by considering the minimum of mean absolute percentage error (MAPE). In their research they found out that ARIMA Model for forecasting farm price of oil palm was ARIMA (2,1,0), while ARIMA Model for forecasting wholesale price of oil palm was ARIMA (1,0,1) or ARMA(1,1), and ARIMA Model for forecasting pure oil price of oil palm was ARIMA (3,0,0) or AR(3).

Javier Contreras, Rosario Espínola, Francisco J. Nogales, and Antonio J. Conejo of the IEEE developed ARIMA models to predict next-day electricity prices in Spain (2003). They had observed that both for spot markets and long-term contracts, price forecasts had become necessary to develop bidding strategies or negotiation skills in order to maximize benefits. This made them to come up with a paper that provided a method to predict next-day electricity prices based on the ARIMA methodology. ARIMA techniques were used to analyze time series. The applications of time-series analysis for computer networks (2006) were done by Marcos Portnoi and Martin Swany of the department of computer and Information Sciences in the University of Delaware. In their work, applications for various time-series statistical analysis models of computer network data, specifically for performance and anomaly detection was done. They said that for the provision of quality services, network agents must control the fair distribution of resources based on historical behavior of applications, rather than on deterministic algorithms, i.e. Virtual circuits could be allocated on demand for applications that exhibit a past of high utilization. Furthermore, in a network performance monitoring architecture, such as perfSONAR, services could benefit from time-series analysis of measurement data to trigger events, audit statistical behavior, or detect anomalies in the network. These anomalities could indicate performance or security issues. They concluded their study by stating that, time-series could enable forecasting that could be employed to predict future performance of the network.

2.3 CONCLUSION

In all these research work carried out on influx of refugees by various researchers, it can be noted that no research was carried out to extrapolate the number of refugees. Similarly, in the studies of time series models, ARIMA model was not used to forecast the number of people. This is why in this research ARIMA model of time series is used to forecast the number of refugees.

III. RESEARCH METHODOLOGY

3.1 INTRODUCTION

Auto Regressive Integrated Moving Average (ARIMA) model was introduced by Box and Jenkins (thus called Box Jenkins model) in 1960s for forecasting. In this project ARIMA model for influx of number of refugees in Kenya is developed. Since ARIMA method is an extrapolation method for forecasting, it only requires the historical time series data on the variable under forecasting. It is one of the most sophisticated methods in extrapolation since it incorporates the features of all such methods, it does not require one to choose the initial values of any variable and values of various parameters prior, and it is robust to handle any data pattern. But this model is difficult to develop and apply in that it involves transformation of the variable, identification of the model, estimation through non linear method, verification of the model and derivation of forecasts. In this chapter, first the ARIMA model is explained, and then we developed the same for the influx of refugees in Kenya from the year 1993 to 2010, and finally applied the same in forecasting the values of the variable for some years to come.

3.2 THEORETICAL BASIS OF TIME SERIES ANALYSIS

Time series is a set of repeated observations of the same variable over a certain period of time. The term time series comes from econometrics studies in which the index variable refers to intervals of time measured in a suitable scale.

The p[th] order autoregressive component relates each value $Z_t = Y_t - (trend\ and\ seasonality)$ to the p, Z_t previous values. Hence the AR(p), has a general form of

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots\ldots + \phi_p z_{t-p} + \varepsilon_t \qquad (3.0.1)$$

Where

Z_t = number of refugees in a given year

$z_{t-1}, z_{t-2}, \ldots\ldots z_{t-p}$ = number of refugees from the previous years

$\theta_1, \theta_2, \ldots\ldots\ldots \ldots \theta_p$ =Coefficients to be estimated

ε_t = error in number of refugees in a given year

A qth-order moving average component: MA (q) has a general form of;

$$z_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots\ldots - \theta_q \varepsilon_{t-q} \qquad (3.0.2)$$

Where

Z_t = number of refugees in a given year

μ = constant mean of refugees

$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\ldots \varepsilon_{t-q}$ = error in number of refugees from given years

$\theta_1, \theta_2, \ldots\ldots\ldots \ldots \theta_q$ = coefficients to be estimated

The p[th] order autoregressive model and the q[th] order moving average model are then combined to form Autoregressive Moving Average Model, ARMA (p, q), which has the form,

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots\ldots \phi_p z_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots\ldots - \theta_q \varepsilon_{t-q} \qquad (3.0.3)$$

At this stage, the graph of sample autocorrelation function (ACF) and partial autocorrelation function (PACF) is drawn. The ACF ρ (k) at lag k of the z_t series is the linear correlation coefficient between Z_t and Z_{t-k} calculated for k =0, 1, 2... by

$$\rho(k) = \frac{\operatorname{cov}(z_t z_{t-k})}{\sqrt{\operatorname{var}(z_t)\operatorname{var}(z_{t-k})}} \qquad (3.0.4)$$

while the PACF is defined as the linear correlation between Z_t and Z_{t-k}, controlling for possible effects of linear relationships among values at intermediate lags, also calculated by equation 3.0.4 above. The ACF and PACF graphs are important in determining the model to be used, which can be summarized as shown in the table 3.1 below

Table 3.1

Model	**ACF**	**PACF**
AR(p)	Dies down	Cut off after lag q
MA(q)	Cut off after lag p	Dies down
ARMA(p q)	Dies down	Dies down

If a graph of ACF of the time series values either cuts off fairly quickly or dies down fairly quickly, then the time series values are considered stationary. If a graph of ACF dies down extremely slowly, then the time series values are considered non-stationary If the series is found to be not stationary, it is converted to stationary series by differencing. That is, the original series is replaced by a series of differences. An ARMA model is then specified for the differenced series. Differencing is done until a plot of the data indicates the series varies about a fixed level, and the graph of ACF either cuts off fairly quickly or dies down fairly quickly.

In the theory of time-series analysis a highly useful operator is the lag or backward linear operator (B) defined by

$$BY_t = Y_{t-1} \quad (3.0.5)$$

The Autoregressive integrated moving average model is denoted by ARIMA (p, d, q), where p indicates the order of the autoregressive part, d indicates the amount of differencing, and q indicates the order of the moving average part. If the original series is found to be stationary, d = 0 and the ARIMA models reduce to the ARMA models, the difference linear operator (Δ), will be defined by

$$\Delta Y_t = (1 - B)Y_t \quad (3.0.6)$$

The stationary series Wt obtained as the d[th] difference (Δd) of Y_t ,

$$W_t = \Delta^d Y_t = (1 - B)^d Y_t \quad (3.0.7)$$

The ARIMA (p d q) model will have the general form of

$$\phi_p B(1 - B)^d Y_t = \mu + \theta_q \varepsilon_t \quad (3.0.8)$$

Once the stationary is obtained, the theory in the table above helps in identification of the model to be used. The model is then checked for adequacy by considering the properties of the residuals, i.e. whether the residuals from an ARIMA model has the normal distribution and is random. The overall check of model adequacy is provided by the Ljung-Box Q statistic. The test statistic Q is given by

$$Q_m = n(n + 2)\sum_{k=1}^{m} \frac{r_k^2(e)}{n - k} \approx \chi^2_{m-r} \quad (3.0.9)$$

Where rk(e) = the residual autocorrelation at lag k

n= the number of residuals

m= the number of time lags included in the test

If the p-value associated with the Q statistic will be small (p-value < α), the model will be considered inadequate. The new model will be considered or the model will be modified and continue the analysis until a satisfactory model will be determined The normality check will be done by considering the p-value from one-sample Kolmogorov-Smirnov test. The randomness of the residuals will be checked by considering the graph of ACF and PACF of the residual. The individual residual autocorrelation should be small and generally within ±2/√n of zero.

The forecasting equation for j years ahead will be given as

$$z_{t+j} = \phi_0 + \phi_1 z_{t+j-1} + \phi_2 z_{t+j-2} + \ldots\ldots \phi_p z_{t+j-p} - \theta_1 \varepsilon_{t+j-1} - \theta_2 \varepsilon_{t+j-2} - \ldots\ldots - \theta_q \varepsilon_{t+j-q} \quad (3.1.10)$$

IV. DATA ANALYSIS AND PRESENTATION

4.1 BUILDING ARIMA MODEL FOR INFLUX OF REFUGEES DATA.

Year	Number of refugees
1993	301,595
1994	253,423
1995	234,665
1996	223,640
1997	232,097
1998	238,197
1999	223,696
2000	218,510
2001	250,960
2002	220,930
2003	241,641
2004	249,310
2005	267,731
2006	390,900
2007	371,630
2008	1,179,990
2009	882,330
2010	985,260

Table 1: Number of refugees

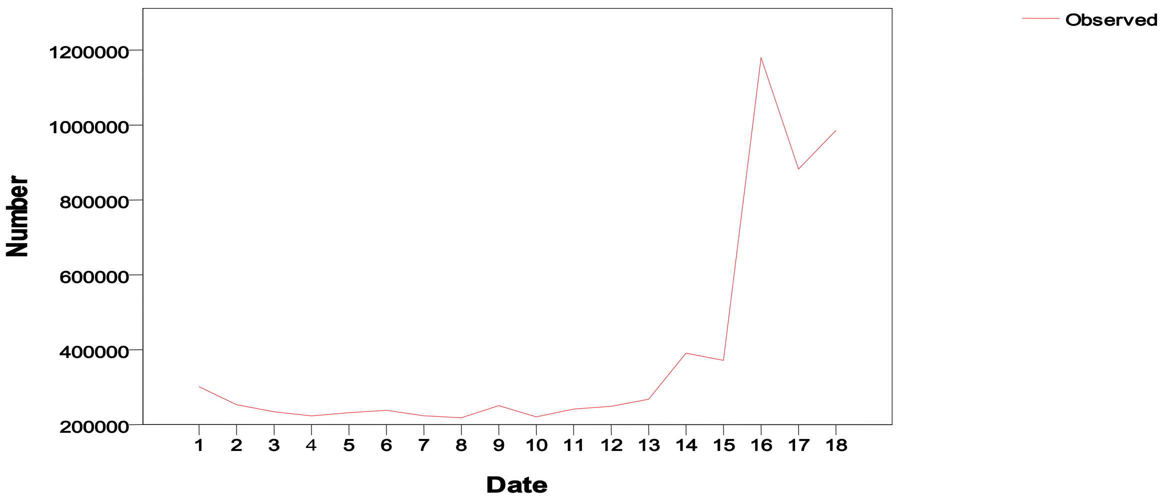


Figure1: Data from 1993 to 2010

In this study, the data for influx of refugees is for a period 1993 to 2010.The development of ARIMA model involves three steps model identification, estimation and verification. The steps are explained for influx of refugees.

4.2 MODEL IDENTIFICATION

ARIMA model was estimated only after transforming the variable under forecasting i.e. number of refugees, into a stationary series. To check for stationary, we examined the time plot of the data. It was revealed through graph 1 that the data is non stationary. We converted the collected data to stationary by differencing. In this case difference of order 1 was sufficient to achieve stationary in mean. The newly constructed variable X_t was then examined for stationary. The graph of new X_t was found to be stationary in mean. The next step was identification of the values of p and q. For this, the autocorrelation (table 2) and partial autocorrelation (table 3) coefficients of various orders of X_t are computed.

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	-.421	.223	3.576	1	.059
2	.199	.215	4.428	2	.109
3	-.043	.208	4.472	3	.215
4	-.004	.201	4.472	4	.346
5	.016	.193	4.479	5	.483
6	-.056	.185	4.570	6	.600
7	.030	.176	4.599	7	.709
8	-.036	.167	4.645	8	.795
9	-.031	.157	4.683	9	.861
10	-.012	.147	4.689	10	.911
11	-.022	.136	4.715	11	.944
12	-.046	.124	4.853	12	.963
13	-.033	.111	4.942	13	.976
14	-.069	.096	5.448	14	.979
15	.035	.079	5.646	15	.985

Table 2: Autocorrelation coefficients

Partial Autocorrelations

Lag	Partial Autocorrelation	Std. Error
1	-.421	.243
2	.026	.243
3	.060	.243
4	-.009	.243
5	.002	.243
6	-.054	.243
7	-.015	.243
8	-.020	.243
9	-.061	.243
10	-.052	.243
11	-.038	.243
12	-.077	.243
13	-.090	.243
14	-.128	.243
15	-.037	.243

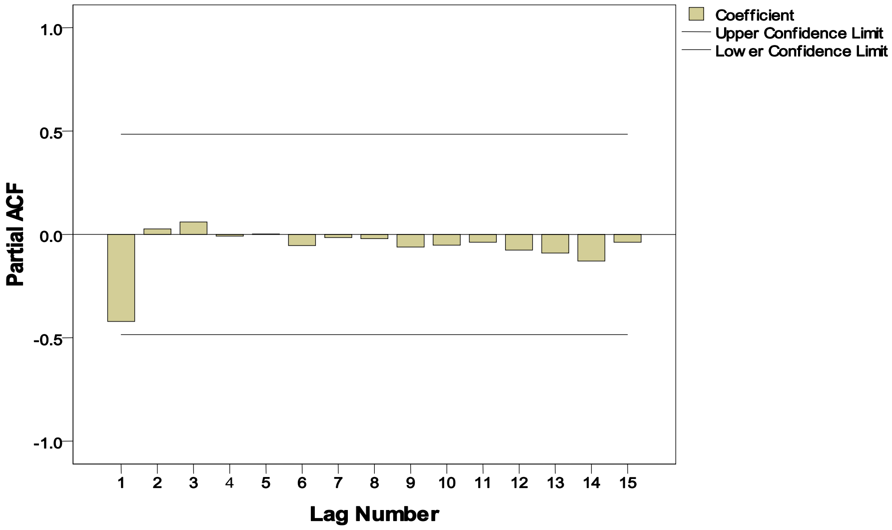
Table 3: Partial Autocorrelation function

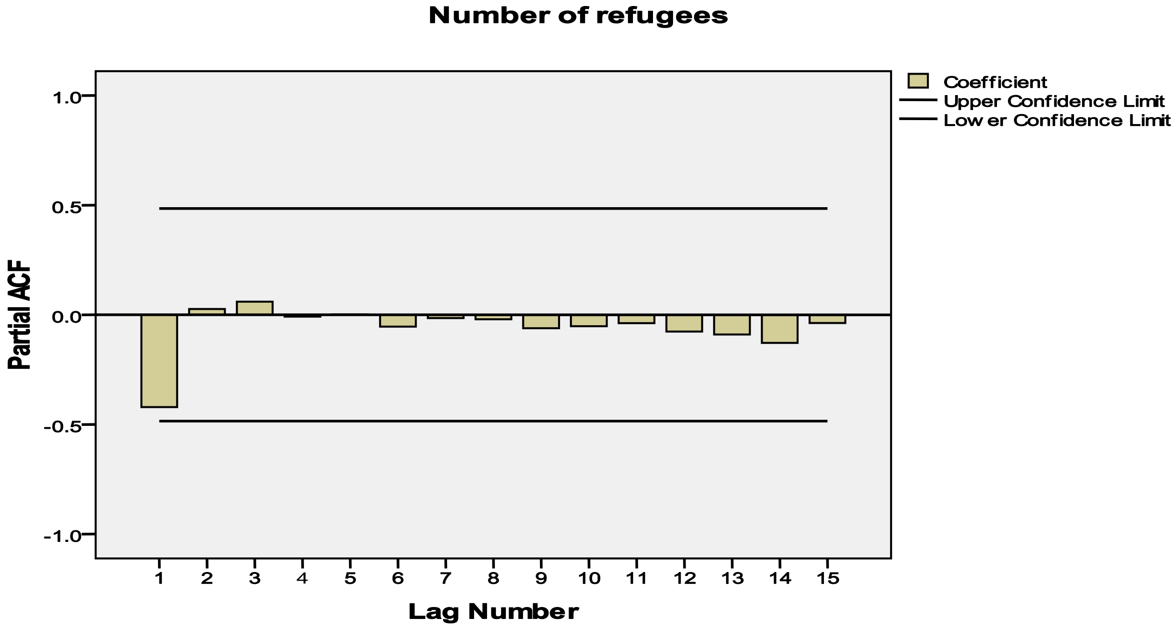
The ACF (graph 2) and PACF (graph 3) showed that the order of p and q could be at most 2.

We examined two tentative ARIMA models and chose the model which had minimum BIC (Bayesian Information Criterion). The models and corresponding BIC values were;

ARIMA (p, d, q)	Normalized BIC
210	24.925
211	25.002

So the most suitable model was found to be ARIMA (2, 1, 0) since this model had the lowest BIC value.





Graph 2: ACF and PACF of Residuals

4.3 ESTIMATION AND VERIFICATION OF THE MODEL

Results of estimation were recorded as shown in table 4.

Table 4: ARIMA Model Parameters

			Estimate	SE	t	Sig.
Number of refugees	Constant		-2.788E6	1.034E6	-2.698	.018
	AR	Lag 1	-.720	.298	-2.415	.031
		Lag 2	-.404	.366	-1.104	.289
	Difference		1			
Year	Numerator	Lag 0	13951.147	5163.221	2.702	.018

Model verification was concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen ARIMA

This was done through examining the autocorrelations and partial autocorrelations of the residuals of various orders as shown in figure 4. For this purpose, the various correlations up to 16 lags were computed and the same along with their significance which was tested by Box-Ljung test as provided in table 2. As the results indicate, none of these correlations were significantly different from zero at a reasonable level. This proved that the selected ARIMA model is an appropriate model. The ACF and PACF of the residuals (fig 4 and 5) also indicated 'good fit' of the model. So the fitted ARIMA model for the number of refugee data was

$$Z_t = 2.788E6 - 0.720 Z_{t-1} + 0.404 Z_{t-2} + \varepsilon_t \qquad 4.0.2.1$$

In the analysis of the ACF and PACF of residuals, it can be noted that ACF and PACF values at lag k >:0 are not statistically different from zero. This proved that the chosen model is a good fit model.

4.4 FORECASTING USING ARIMA MODEL

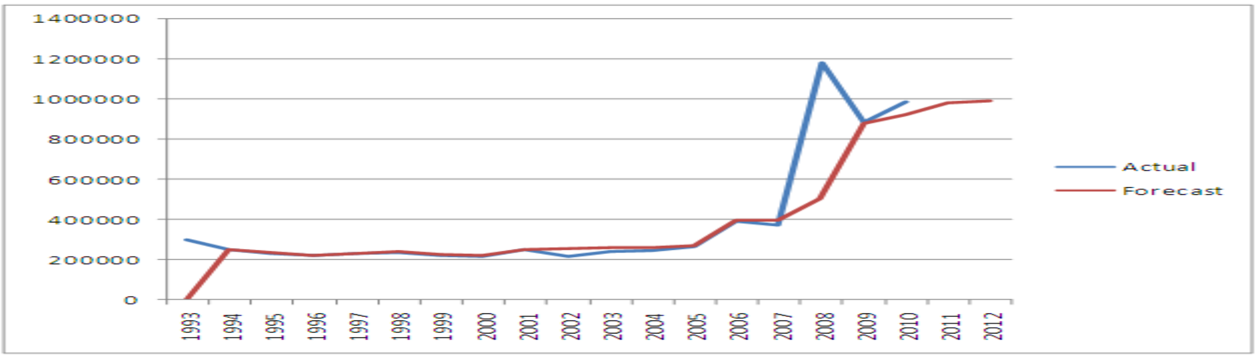
With the help of forecasting equation, the developed ARIMA model was used in forecasting the number of refugees. The two types of forecasting were; sample period forecasts and post-sample period forecasts.

4.4.1 SAMPLE PERIOD FORECASTS

The sample period forecasts of the number of refugees were found and recorded.. The forecasted values and the actual values of Zt are shown in table 6.

Year	Actual value	Forecast value	Error
1993	301,595	_	-
1994	253,423	250,856	-2,567
1995	234,665	232,438	-2,227
1996	223,640	221,989	-2,202
1997	232,097	230,267	-1,830
1998	238,197	237,576	-621
1999	223,696	223,696	0
2000	218,510	220,587	2,077
2001	250,960	251,678	718
2002	220,930	256,367	35,437
2003	241,641	257,196	15,555
2004	249,310	260,924	11,614
2005	267,731	270,705	2,974
2006	390,900	395,534	4,634
2007	371,630	396,279	24,649
2008	1,179,990	502,704	-677,286
2009	882,330	882,330	0
2010	985,260	925,167	-60,093
2011	-	982,357	-
2012	-	992,648	-

Table 6: Forecasted and Actual values.



Graph 6: Actual and Forecasted values

4.4.2 POST- SAMPLE PERIOD FORECASTS

The forecasting ability of the fitted ARIMA model is examined by measuring the accuracy of sample period forecasts. The Mean Absolute Percentage Error (MAPE) for number of refugees turns out to be 24.532. This measure indicates that the forecasting inaccuracy is low; hence the forecasted casted values had low level of inaccuracy.

Fit Statistic	Mean	Minimum	Maximum
Stationary R-squared	.403	.403	.403
R-squared	.703	.703	.703
RMSE	185165.014	185165.014	185165.014
MAPE	24.532	24.532	24.532
MaxAPE	50.174	50.174	50.174
MAE	95680.062	95680.062	95680.062
MaxAE	590189.079	590189.079	590189.079
Normalized BIC	24.925	24.925	24.925

4.5 TEST FOR NORMALITY

By the use of Kolmogorov-Smirnov test, we find that the calculated value of test statistic D is 3.07, and the tabulated value of test statistic D is 0.34. Since Dtabulated < Dcalculated, we conclude that the collected data does not follow a normal distribution.

V. CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

ARIMA model offers a good technique for predicting the magnitude of any variable.

The strength of ARIMA model lies in the fact that this method is suitable for any time series with any pattern of change and it does not require the forecaster to choose a prior value of any parameter. However, it requires a long time series data. Like any other method, this technique also does not guarantee perfect forecasts. Nevertheless, it can be successfully used for forecasting long time series data. In our study ARIMA (2,1,0) model was best suited for estimation of influx of refugee data. From the forecast values it can be noted that the trend is upward.

5.2 RECOMMENDATIONS

The model can be used by the ministry of immigration and UNHCR in forecasting the number of refugees in Kenya. However, this model should be updated concurrently with time so as to incorporate current data.

ACKNOWLEDGEMENTS

We would like to acknowledge Patrick Kamau, Zipporah Sinino and Nicholas Muthama for their tireless efforts in carrying out this research and their advice on the technical aspects of data analysis and modeling. May God bless you abundantly.

REFERENCES

[1] Hamilton, J. D. (1994): 'Time Series Analysis', Princeton University Press.
[2] Time Series Analysis: Univariate and Multivariate methods by William W.S. Wei, 1994, pg 208-245
[3] Time Series: Theory and methods by Brockwell Peter J. and Davis, Richard A (1991), Second Edition, Pg 288-305.
[4] Introduction to Time Series and Forecasting by Peter J. Brockwell, Richard A. Davis (1993), Pg 167-140.
[5] B. N. Mandal 'Forecasting sugarcane production in India with ARIMA model'
[6] Javier Contier 'ARIMA Models to Predict Next-Day Electricity Price, IEEE Transactions on Power Systems Vol 18; No. 3, August, 2003.
[7] Ministry of Immigration, Department of Refugees, Kenya.

[8] Najeeb Iqbal, Khuda Bakhsh1, Asif Maqbool And Abid Shohab Ahmad; 'Use of the ARIMA Model for Forecasting Wheat Area and Production in Pakistan'
[9] Rachana Wankhade1, Suvarna Mahalle2, Sonal Gajbhiye and V.M. Bodad. 'Use of the ARIMA Model for Forecasting Pigeon Pea Production in India'
[10] UNHCR statistics archives.
[11] Box, G.E.P., and G. M. Jenkins. 1970. Time series analysis: forecasting and control. pg 205-218
[12] Brockwell, P.J., and Davis, R. A. 1996. Introduction to time series and forecasting. Springer, pg48-55.
[13] Kendall, M. G., and A. Stuart. 1966. The advanced theory of statistics. Vol. 3. Design and Analysis and Time-Series pg 56-76. Charles Griffin & Co. Ltd., London, UK.

Biographies and Photographs

Mr. Allan Lenard Mudanyi completed his Bsc Mathematics and Computer Science from Taita Taveta University College in 2011. Currently he is working with national bureau of statistics.

Mr. Nicholas Muthama Mutua is an Assistant Lecturer and Chairman of the Department of Mathematics and Informatics, Taita Taveta University College, a Constituent College of the Jomo Kenyatta University of Agriculture and Technology. His research interests are mathematical modeling and information systems engineering. He can be reached at: 635-80300. Voi-Kenya. Tel: +254726711097

Mrs. Zipporah Sinino is an Assistant Lecturer in the Department of Mathematics and Informatics, Taita Taveta University College, a Constituent College of the Jomo Kenyatta University of Agriculture and Technology. Her research interests are operations research and business systems modeling. She can be reached at: 635-80300. Voi-Kenya. Tel: +254721814150