Forecasting The Number of International Muslims Umrah Visitors in Makkah to 2030

Ahmad Alamoudi* & Seraj Abed

Department of Industrial Engineering, Faculty of Engineering, King Abdulaziz University
Jeddah, Saudi Arabia,
*Corresponding Author E-mail: *aalamudi0003@stu.kau.edu.sa*

Abstract: This study was conducted to support the Saudi Vision 2030 to move beyond oil dependency by diversifying the sources of its income by increasing the tourists demand in the Kingdom, especially religious tourism which provides more money flowing continuously in the economy depending on its strategic location at the center of the world. This study aims to forecast the annual number of International Umrah visitors to Makkah from the period of 2020 to 2030, to achieve the 2030 Saudi Vision of serving 30 million visitors in 2030. Historical data of international Umrah visitors were collected for the past 33 years on an annual basis. This univariate time series of the annual number of international Umrah visitor has been studied and forecasted to predict the annual number of Umrah visitors for the next 11 years. Different forecasting techniques were used to identify the best technique among them to forecast the future demand. The best technique among those used in the study was ARIMA (2, 2, 2) model as best fitted model with the least residual measures of 9% MAPE. The results of the study indicate that annual number of international Umrah visitors will increase from 7.6 million in 2020 to 12.3 million in 2030. Annual increment percentage decreased from 11% in the historical data to 5% in the forecasted data affected by recent year's growth due to the variability of issuing Umrah visas before 2030 vision's announcement. It is clearly shown based on historical data and recent growth in Umrah visitors the 30 million visitors' goal in 2030 will not be achieved. To achieve the set goal of international Umrah visitors by 2030 the Saudi government would need to increase the growth rate of issuing Umrah visa with an average of 12% approximately. In addition, lots of actions to attract more Umrah visitors such as aggressive marketing campaign, improve its housing & hospitality facilities, increase the capacity of its utility systems, and expand the overall infrastructure in the Holy places.

Keywords: Forecasting; Umrah visitors; Pilgrims; Religious Tourism; Makkah; Saudi Vision 2030

I. INTRODUCTION

The Holy Mosque in Makkah captivates the hearts of Muslims everywhere in the world. It is the focal point (Qiblah) to which 1.8 billion Muslims pray 5 times daily. The Muslims have great desires and dreams to visit the Holy Mosque & places to perform Pilgrim (Hajj) and Visiting (Umrah) as many times as the can if they can afford to do so and capable to do so. Very poor Muslims work for decades, spending their lifetime to collect saving just to visit Makkah and perform Hajj & Umrah, although these hardships. All this happens in response to Abraham's old call after he built the Kaaba with his son Ismail, peace be upon them both.

Pilgrim (Hajj) is the visit to the Holy Mosque & places in Makkah, which is required once in a lifetime, by a capable Muslim during the second week of the twelfth month in Hijri (lunar) calendar. Contrariwise, visiting (Umrah) is visiting only the Holy Mosque with specific rituals by Prophet

Mohammed pace be upon him. Both pilgrims & Umrah visitors called the guest of Allah (God) and they are treated from government and locals with all kindness and welcoming.

On April 25, 2016, Prince Mohammed bin Salman presented Saudi Vision 2030, a sweeping plan to move the kingdom beyond oil dependency. This plan carries a lot of changes to the economy. Saudi Vision 2030 aims to develop the country's ports, cultural resources, and tourist sites to take advantage of its strategic position at the center of the Arab and Islamic world. The first Pillar of the 2030 vision is a vibrant society which will focus the efforts to serve Umrah visitors by taking pride in national identity and living by Islamic values. The Vision was cascaded into strategic objectives to enable effective implementation. Also, these strategic Objectives were cascaded based on a clear methodology includes objectives' setting and overlap analysis, objectives cascaded to 3 levels to ensure inclusiveness and finally identification of metrics and targets to be achieved.

To translate objectives to strategic action plans, Vision Realization Programs (VRPs) were developed to deliver against strategic objectives. The VRP which is interested in achieving the first objective strengthen Islamic values & national identity is enriching the Hajj and Umrah Experience Program [1]. Enriching the Hajj and Umrah Experience Program aims to increase served Umrah visitors from around 7.5 million currently to 30 million visitors annually.

This study aims to forecast annual number of Umrah visitors to Makkah & Madinah from 2020 to 2030 to achieve Saudi Vision 2030 of serving 30 million visitors.

II. LITERATURE REVIEW

Good planning is essential for any organization and seeking success requires an accurate forecasting system can be relied on it. There is no doubt that the organization of Hajj and Umrah that include many interlocking phenomena together needs to have such a system in order to manage the organization in a proven scientific way that achieves the desired goals. On top of these phenomena are both numbers of pilgrims and Umrah visitors themselves, a good forecasting for their numbers is essential for the planning of various important aspects, such as: transportation, housing, communication, etc. Forecasting plays a significant role in Hajj & Umrah planning and it is crucial for Saudi Arabia & the private sector to avoid shortages or surpluses in goods and services.

Published research discussing Umrah's visitors forecasting are very rare, the researchers usually work on economic and statistical studies on Umrah visitors not forecasting. One example of forecasting Umrah visitors' number is Alrekabi, Mukhles [2], the study provided good forecasting for the number of Umrah visitors coming from outside Saudi Arabia for the following 24 months based on the actual numbers of Umrah visitors coming from abroad (out of Saudi Arabia) using Box–Jenkins methodology (ARIMA) models. The researcher applied many transformations of the data such as differencing algorithms and square root to find the best fitting models.

Alkalhout, Jamal [3] this study tried to solve the most important question in terms of planning and decision-making for Umrah seasons, what is the total number of Umrah performers coming from inside and outside the Kingdom. The real problem here is to estimate the number of Umrah performers inside the Kingdom as there is no effective way currently to count their numbers. Unlike the outside Umrah performers who get a visa called Umrah visa, the Umrah performers from inside have no restrictions to perform Umrah any time. The study revealed many important estimates after surveying 21,376 visitors were from outside & inside the Kingdom 68.4% & 31.6% respectively.

According to Song, Witt & Li [4] in their famous book The Advanced Econometrics of Tourism Demand "Tourism visits can take place for various reasons: holidays, business trips, visits to friends and relatives (VFR), conferences, pilgrimages and so on.", therefore we will classify Hajj & Umrah forecast modeling & techniques as tourism demand forecast. "Tourism demand forecasting is a well-established research area, and it has attracted many studies in the tourism and hospitality field.

Burger, Dohnal, Kathrada & Law [5] this paper compared a variety of time-series forecasting methods to predict tourism demand for a certain region, the research has attempted to examine the forecasting accuracy of various time-series tourism forecasting models. Tourist arrivals from the United States to Durban in South Africa were used for model calibration and testing. Many forecasting techniques are employed in this survey such as moving average, decomposition, single exponential smoothing, ARIMA, multiple regression, genetic regression and neural networks, etc. The survey shows that the neural network method performs the best by achieving the lowest MAPE.

Hassani, Silva, Antonakakis, Filis & Gupta [6] this paper aimed to evaluate the use of several parametric and nonparametric forecasting techniques for predicting tourism demand in selected European countries. As many other papers were reviewed in this literature, they found that, no single model can provide the best forecasts for any of the countries in the short, medium and long-run.

III. METHODOLOGY

The methodology of this study was to apply the forecasting steps starting from problem formulation till implementation of the forecast. The steps are briefly explained below and shown in Fig. 1

A. Problem Formulation:

The demand of Umrah visitors historically changes over time affecting many aspects of Saudi Arabia preparations to serve the increasing number of Umrah visitors. The problem under study is developing an excellent forecasting model that can predict outside Umrah performers from 2020 to 2030.

B. Data Collection:

Different ministries & authorities had been visited or contacted to collect the data such as Ministry of Hajj and Umrah, Ministry of Tourism, General Authority for Statistics, and the Custodian of Two Holy Mosques Institute for Hajj and Umrah research in Umm Al-Qura University to collect the data.

C. Data Manipulation & Cleaning:

Autocorrelation Function (ACF): is the correlation between a variable lagged one or more time periods and itself. Data patterns, including components such as trend and seasonality, can be studied using autocorrelations. Hanke & Wichern [7]. Based on the techniques used in forecasting different manipulations has been done on the data such as such as differencing and logarithms to find the best fitting models.

D. Model Building:

Choosing the best recommended forecasting techniques based on the data patterns appear after autocorrelation has been done.

E. Model Evaluation:

A residual is the difference between an actual observed value and its forecast value. To select the best fitted forecasting techniques among the recommended ones. The best model obviously is the one with

the least forecasting error. Three residuals measures picked out to compare the results of forecasting techniques which are:

- Mean Absolute Deviation (MAD)
- Mean Square Error (MSE)
- Mean Absolute Percentage Error (MAPE)

F. Model Implementation & Forecast Evaluation:

Applying the best fitted forecasting technique to find the forecast Umrah visitors for the period of (1987-2019) and evaluate the results by taking the residuals' randomness.

G. Forecast Implementation and Feedback & Updating:

Generating the forecast for the future period of 2020-2030 by the best model. Then, once any given year from 2020 to 2030 end the actual number of Umrah visitors should be updated and revised to measure the error in this forecast and to update the model to generate another version of the forecast.

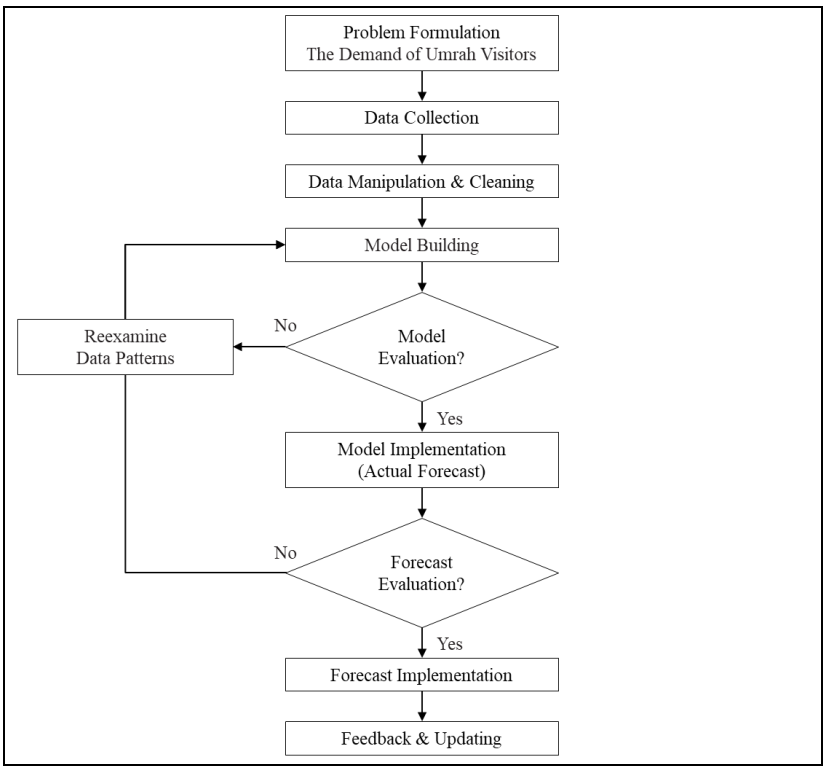

Figure 1. Forecasting Methodology

IV. RESULTS

The time series in this study is a univariate time series,

A. Data Collection:

All the data used in this study was gathered from different sources rely on two main sources. Number of outside Umrah visitors (abroad pilgrims) is essentially defined by the number of Umrah visas determined by Ministry of Hajj & Umrah and issued by Minister of Interior, the General Directorate of Passports. Data has been collected for the period (1987-2019) representing 33 years. The available, valid and reliable data is for those visitors coming outside Saudi Arabia because they must get Umrah visa to pass Saudi Arabia boarders.

B. Data Manipulation & Cleaning:

Stationarity Tests:

1. Graphical Analysis:

We can examine whether a series is stationary or not by analyzing the time plot of that particular series. In this regard, a time plot of the outside Umrah visitor's series is shown in Fig. 2 below:

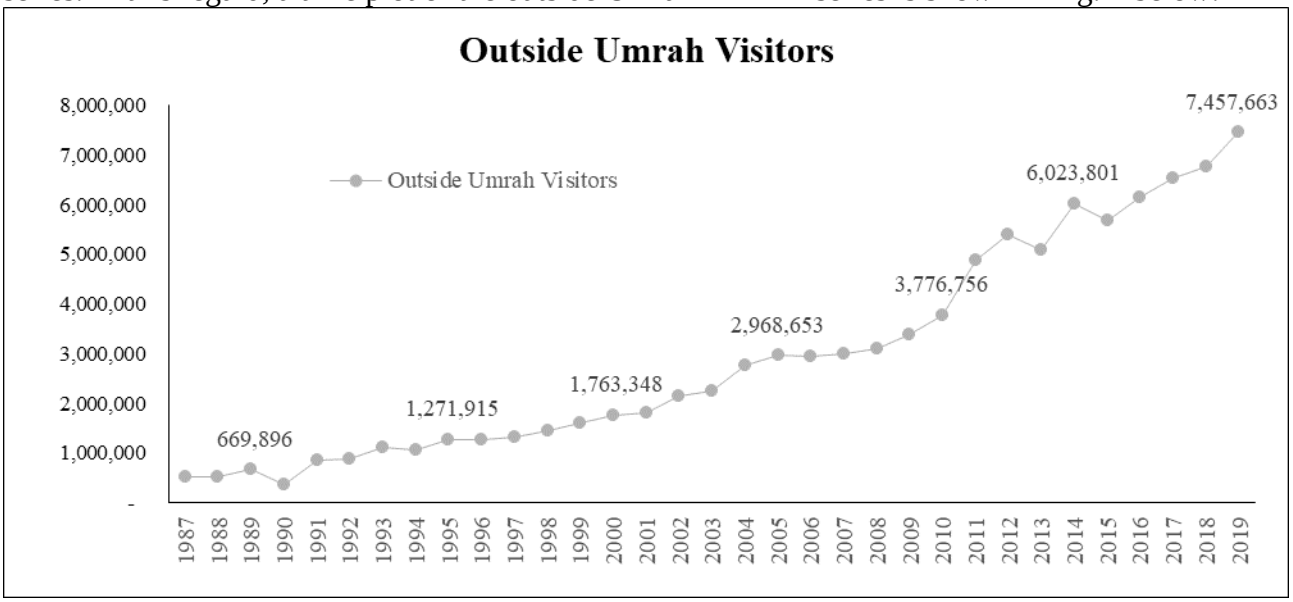

Figure 2. Number of Outside Umrah Visitors **1987 – 2019.**

The above graph shows that outside Umrah visitors' series is not stationary since it is trending upwards over the period under study. The implication is that the mean of series is changing over time and hence we can safely conclude that the variance of series is not constant over time.

2. Autocorrelation (Correlogram) Analysis:

The correlogram below confirms our analysis from the observation of the time series plot of outside Umrah visitors. ACF coefficients are very high particularly for the first 3 lags and is very typical in non – stationary time series data. Partial Autocorrelation Function (PACF) also plotted below.

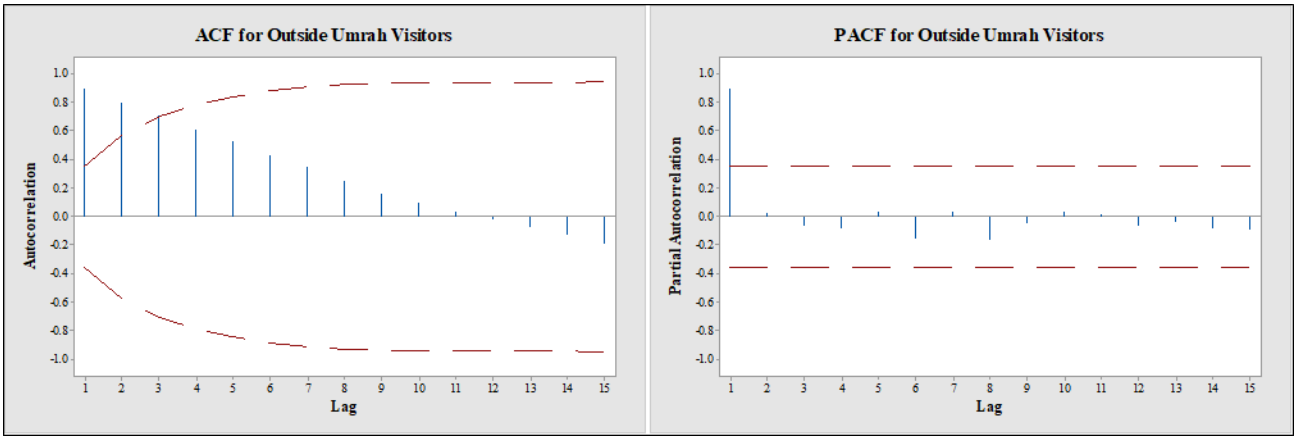

Figure 3. Correlogram ACF & PACF for outside Umrah visitors (15 lags)

The truly random series has ACF coefficients that should lie within a range specified by zero, plus or minus a certain number of standard errors. At a 95% confidence level a series can be considered random if each of the calculated ACF coefficients is within the interval about 0 given by 0 ± t × SE(r_k).

$$SE(r_k) = \sqrt{\frac{1+2\sum_{i=1}^{k-1} r_i^2}{n}} \quad \textbf{(1)}$$

Where,
$SE(r_k) =$ the standard error (estimated standard deviation) of the autocorrelation at time lag k
r_i = the autocorrelation at time lag i
k = the time lag
n = the number of observations in the time series.

C. Model Building:

According to Hanke & Wichern [7], there are a number of suggested techniques to deal with the trending upwards univariate time sires include moving averages, Holt's linear exponential smoothing, simple regression, growth curves, exponential models, and autoregressive integrated moving average (ARIMA) models (Box-Jenkins methods).

D. Model Evaluation:

Minitab® 17.1.0 computer package used to compute forecasting residual measures for each forecasting method & model then summarized in table 1. The methods were ranked in ascending order starting from the lowest in all methods and the best forecasting methods among them all is **ARIMA (2,2,2):**

- Mean Absolute Deviation (MAD) = **180,908**
- Mean Square Error (MSE) = **72,153,531,464**
- Mean Absolute Percentage Error (MAPE) = **9%**

Table 1. shows summary of forecasting error measures for different forecasting techniques:

Method	*Model*	*MAD*	*MSE*	*MAPE (%)*
ARIMA	**(2,2,2)**	180,908	72,153,531,464	9
ARIMA	**(1,2,2)**	206,289	89,447,586,059	11
Holt's Linear Exponential Smoothing		216,749	92,339,600,000	13
Single Exponential Smoothing		267,964	138,744,000,000	12
Moving Averages	**MA 2**	341,650	207,158,000,000	14
Double Moving Averages	**DMA 2,2**	445,119	317,970,817,345	17
Moving Averages	**MA 3**	445,230	316,579,000,000	17
Simple Linear Regression		511,563	335,436,000,000	32

Fig. 4 shows plots for actual data vs. the best four fitted forecasting techniques ARIMA(2,2,2), ARIMA(1,2,2), Holt's Linear Exponential Smoothing & Single Exponential Smoothing:

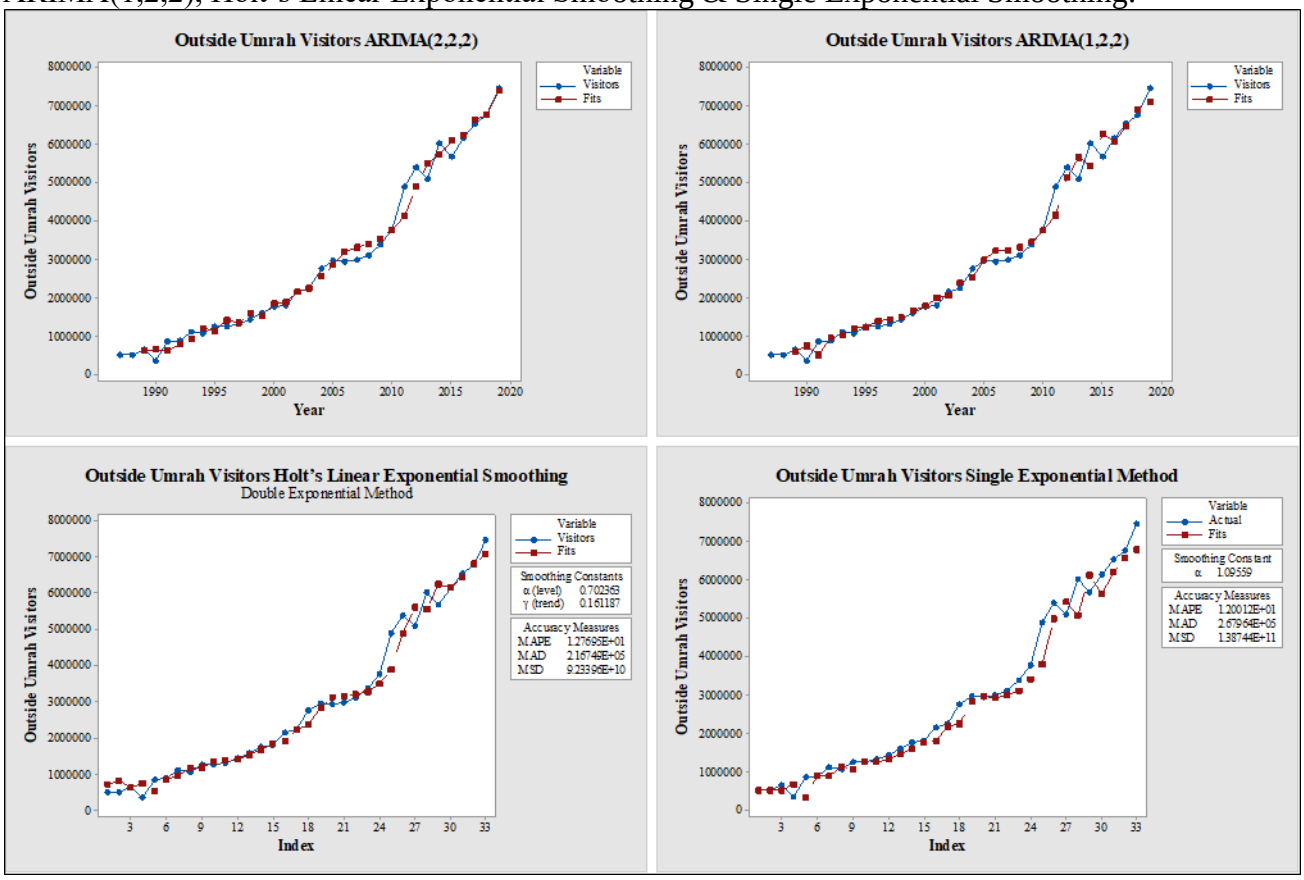

Figure 4. Actual data vs. the best four fitted forecasting techniques

*E. **Model Implementation & Forecast Evaluation:***

1- Model building & estimation procedures:

- Autoregressive Models for Outside Umrah Visitors (OUV):

Given:

$$OUV_t = \emptyset_0 + \emptyset_1 OUV_{t-1} + \cdots + \emptyset_p OUV_{t-p} + \varepsilon_t \quad \textbf{(7)}$$

where,

OUV_t = outside Umrah visitors at time t

OUV_{t-1} = are previous period values of the outside Umrah visitors

$\emptyset_0$ = the coefficients to be estimated

ε_t = the error term at time t

We say that equation (7) is an Autoregressive (AR) process of order p and is commonly denoted as AR (p).

- Moving Average Models for Outside Umrah Visitors (OUV):

Given:

$$OUV_t = \mu + \varepsilon_t + \omega_1 \varepsilon_{t-1} + \cdots + -\omega_q \varepsilon_{t-q} \quad \textbf{(8)}$$

Where,

OUV_t = outside Umrah visitors at time t

μ = the constant mean of the process

ω_1, ω_q = the coefficients to be estimated

ε_t = the error term,

$\varepsilon_{t-1}, \varepsilon_{t-q}$ = the errors in previous time periods that, at time t

We say that equation (8) is a Moving Average (MA) process of q^{th}-order, commonly denoted as MA (q).

- The Autoregressive Moving Average (ARMA) model

Box & Jenkins (1970) illustrated that, an ARMA (p, q) process is simply a combination of AR (p) and MA (q) processes. Therefore, combining equations (7) and (8) an ARMA (p, q) can be specified as follows:

$$OUV_t = \emptyset_0 + \emptyset_1 OUV_{t-1} + \cdots + \emptyset_p OUV_{t-p} + \varepsilon_t + \mu + \varepsilon_t + \omega_1 \varepsilon_{t-1} + \cdots + -\omega_q \varepsilon_{t-q} \quad \textbf{(9)}$$

ARMA (p, q) model is like the AR (p) and the MA (q) models, can only be used only for stationary time series data. Outside Umrah visitors time series are non – stationary as discussed above. For this simple reason, ARMA models were not used for describing non – stationary time series.

- Autoregressive Integrated Moving Average (ARIMA) Models (Box-Jenkins Methods):

If the series is not stationary, it can often be converted to a 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. Models for nonstationary series are called autoregressive integrated moving average models and denoted by ARIMA (p,d,q). Here, 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 stationary, then, $d = 0$ and the ARIMA models reduce to the ARMA models. Hanke & Wichern [7].

For the outside Umrah visitors we already knew that the data is trendy, and series is not stationary. An upward trend is very clear in Fig. 2; therefore, the first difference was used to make the series stationary. Fig. 5 shows the first difference plotted data, and the data seems to be varied about a fixed level above the zero and that will lead us to use a constant as the coefficient $\emptyset_0$ which is related to the constant level of the series. This can be supported by the ACF & PACF for the first differenced data. Surprisingly, the first difference data are random data with no significant ACF in any lag of the first 15 lags as shown in Fig. 5. To judge their significance, both the sample autocorrelations and the sample partial autocorrelations are usually compared with $\pm 2/\sqrt{n}$ where n is the number of observations in the time series and usually these limits work well when n is large. This led us to use second difference for outside Umrah visitors, here d = 2 and ACF & PACF suggested many models to be tested comparing the theoretical ACF & PACF for the various ARIMA models such AR (1), AR(2), MA(1) & MA(2). AFC for second difference dies out after the first lag and that indicate AR (1) model, PACF for second difference dies out after second lag and that's indicate MA(1) & MA(2) models. Both ACF & PACF die out immediately after the first or second lag and that guide us to use ARMA models for the second difference as shown in third line Fig. 5.

The identification step ends with identifying a group of possible models to go to the next step model estimation and the models that will be estimated are:

- ARIMA(1,2,1)
- ARIMA(2,2,1)

- ARIMA(1,2,2)
- ARIMA(2,2,2)

Figure 5. Shows 1st & 2nd differenced data with ACF & PACF for both

2- Model Checking

Adequacy check must be done before forecast. Basically, a model is adequate if the residuals cannot be used to improve the forecasts. That is, the residuals should be random. The individual residual autocorrelations should be small and generally within $\pm 2/\sqrt{n}$ of zero. Significant residual autocorrelations at low lags or seasonal lags suggest that the model is inadequate and that a new or modified model should be selected.

Fig. 6 shows ACF for the suggested models' residuals. Two of them ARIMA(1,2,1) & ARIMA(2,2,1) have a significant correlation at lag 4 that's suggest that the models are inadequate. Contrariwise ARIMA(2,2,2) & ARIMA(1,2,2) models are showing random residuals with no significant correlation at any lags and this will reduce the number of models to two out of four.
Model checking step ends with identifying suitable models to go to the next step forecasting with the model and the models that will be used to forecast are:

- ARIMA(1,2,2)
- ARIMA(2,2,2)

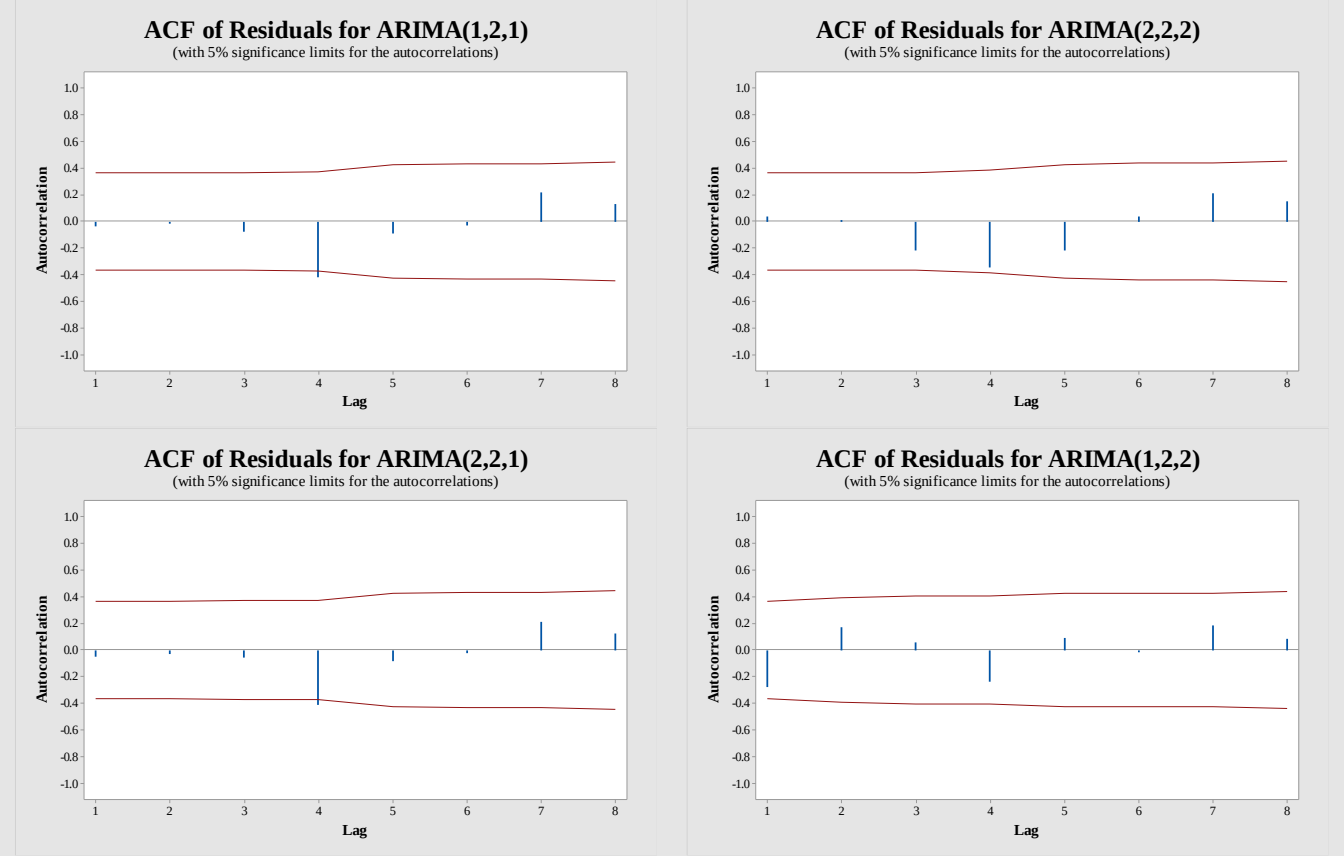

Figure 6. ACF for suggested ARIMA models

F. Forecast Implementation and Feedback & Updating:

Results Presentation for ARIMA(2,2,2):

Table 3 Final Estimates of Parameters:

Variable Type	*Coef*	*SE Coef*	*T*	*P*
AR(1)	-1.3363	0.1886	-7.08	0
AR(2)	-0.4742	0.1901	-2.49	0.019
MA(1)	0.1626	0.218	0.75	0.462
MA(2)	1.0742	0.2182	4.92	0
Constant	29630.1	8.7	3410.18	0

The forecasting for the period from 2020 to 2030 via ARIMA(2,2,2) is shown below in table 4 with upper and lower limits of 95%. A plot of the forecasted 11 years has been added to the actual series as

shown in Fig. 7. The model is showing random residuals with no significant correlation at any lags for ACF & PACF indicating an adequate model to be used in forecasting.

Table 4 shows forecasted number of outside Umrah visitors via ARIMA(2,2,2) Model:

Period (Year)	*Forecast*	*Lower Limit 95%*	*Upper Limit 95%*
2020	7,628,163	7,101,574	8,154,753
2021	8,214,250	7,625,232	8,803,267
2022	8,521,950	7,885,716	9,158,185
2023	9,034,207	8,342,408	9,726,006
2024	9,434,763	8,733,230	10,136,297
2025	9,917,210	9,191,850	10,642,571
2026	10,372,828	9,646,222	11,099,434
2027	10,855,094	10,124,608	11,585,581
2028	11,344,103	10,613,159	12,075,048
2029	11,841,095	11,109,494	12,572,697
2030	12,353,853	11,615,324	13,092,381

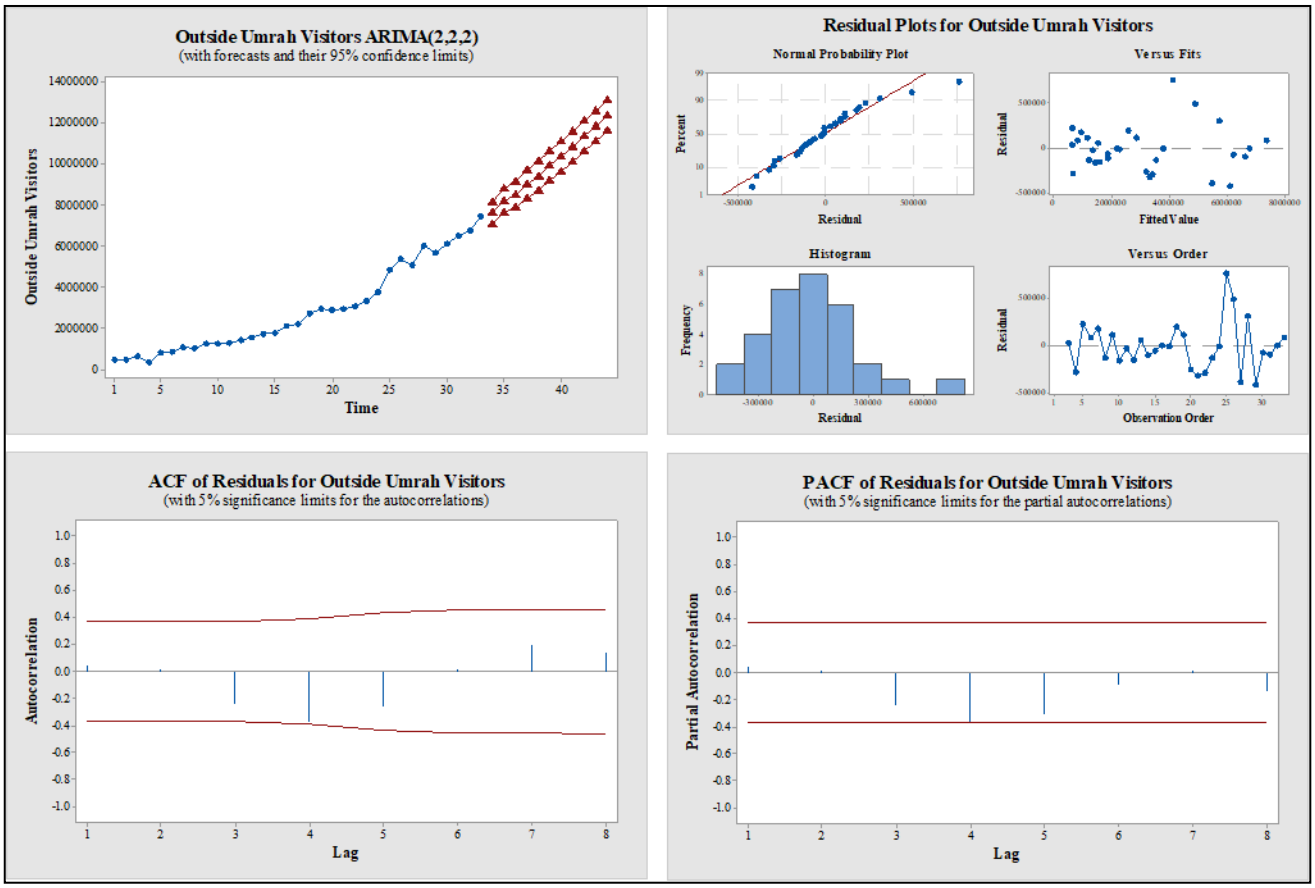

Figure 7. ARIMA(2,2,2) forecasting results, residual plots and residuals ACF & PACF

V. CONCLUSION

The results of the study indicate that annual number of international Umrah visitors will increase from 7.6 million in 2020 to 12.3 million in 2030. These results obtained by using the best technique among those used in the study was ARIMA (2, 2, 2) model, as best fitted model with the least residual measures of 9% MAPE, MAD of 180,908 and MSE of 72,153,531,464. Annual increment percentage decreased from 11% in the historical data to 5% in the forecasted data affected by recent year's growth due to the variability of issuing Umrah visas before 2030 vision's announcement. It is clearly shown based on historical data and recent growth in Umrah visitors the 30 million visitors' goal in 2030 will not be achieved. Moreover, no models reported 30 million visitors in all level of error measures which indicates that by relying on mathematical methods in dealing with predictions, without injecting the judgment or opinion of the analyst or (the concerned authorities), by adding more factors to the forecasting process, it is difficult to achieve Saudi Vision 2030 goal of receiving 30 million pilgrims.

V. RECOMMENDATIONS

To achieve the set goal of international Umrah visitors by 2030 the Saudi government would need to increase the growth rate of issuing Umrah visa with an average of 12% approximately. In addition, lots of actions to attract more Umrah visitors such as aggressive marketing campaign, improve its housing & hospitality facilities, increase the capacity of its utility systems, expand the overall infrastructure in the Holy places, issuing more visas for some Islamic countries with huge population, distributing visitors among Hijri months by issuing more Umrah visas on low season months, plan the incremental number of visitors in advance for the next 10 years.

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical statement: The authors declare that they have followed ethical responsibilities.

REFERENCES

[1] Saudi Vision 2030 (2016). https://vision2030.gov.sa/en

[2] Rekabi, Mukhles (2013) Using Box and Jenkins' methodology in: Predicting the Number of Umrah Pilgrims from Abroad, Every Month. The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research. Study No. 43411001. "unpublished"

[3] Alkalhout, Jamal (2009) Estimating the number of inside Umrah performers in Ramadan 1429 AH and the distribution of pilgrims leaving the doors of the Holy Mosque - Study No. 42914001 "unpublished"

[4] Song, Haiyan & Witt, Stephen & Li, Gang. (2009). The Advanced Econometrics of Tourism Demand. 10.4324/9780203891469.

[5] Burger, C.J.S.C & Dohnal, M. & Kathrada, M & Law, Rob. (2001). A practitioners guide to time-series methods for tourism demand forecasting — a case study of Durban, South Africa. Tourism Management. 22. 403-409. 10.1016/S0261-5177(00)00068-6.

[6] Hassani, Hossein & Silva, Emmanuel & Antonakakis, Nikolaos & Filis, George & Gupta, Rangan. (2017). Forecasting accuracy evaluation of tourist arrivals. Annals of Tourism Research. 63. 112-127. 10.1016/j.annals.2017.01.008.

[7] Hanke, J. E., & Wichern, D. W. (2014). Business forecasting. 9th edition.

[8] Nyoni, T. (2018). Modeling and forecasting inflation in Kenya: Recent insights from ARIMA and GARCH analysis. Dimorian Review, 5(6), 16-40.

[9] General Authority for Statistics. (2020). Umrah survey https://www.stats.gov.sa/en/862

