Modeling and Forecasting the Number of Pilgrims Coming from Outside the Kingdom of Saudi Arabia Using Bayesian and Box-Jenkins Approaches

SAMEER M. SHAARAWY, ESAM A. KHAN, MAHMOUD A. ELGAMAL
The Custodian of the Two Holy Mosques Institute for Hajj and
Omra Research, Umm Al -Qura University
Makkah, Saudi Arabia

***Abstract*—Pilgrimage has received a great attention by the government of Saudi Arabia. Of special interest is the yearly series of the Number of Pilgrims coming from outside the kingdom (NPO) since it is one of the most important indicators in determining the planning mechanism for future hajj seasons. This study approaches the problems of identification, estimation, diagnostic checking and forecasting of the NPO series using Bayesian and Box - Jenkins approaches. The accuracy of Bayesian and Box- Jenkins techniques have been checked for forecasting the future observations and the results were very satisfactory. Moreover, it has been shown that Bayesian technique gives more accurate results than Box-Jenkins technique.**

Keywords—autoregressive processes, identification; estimation; diagnostic checking; forecasting; Jeffreys' prior; and posterior probability mass function.

I. Introduction

Pilgrimage (or hajj) of Moslems is one of the most important events all over the world. It is considered the largest human gathering in which pilgrims move together through a very limited space in a short period of time. This important event is repeated annually, and the number of pilgrims is increasing year after year. In addition, hajj is one of the main sources of gross national product (GNP) in Saudi Arabia. Therefore, it has received a great attention by the government of Saudi Arabia. Every year, the kingdom of Saudi Arabia spends a great deal of efforts and money to improve the hajj system, which includes security, economy, management of water and electrical resources, services and goods required by the vast number of pilgrims. However, without knowing the number of pilgrims in advance may make the process of improvement very difficult. Therefore, it is very important to have a mechanism for predicting and forecasting the number of pilgrims in order to determine the size and quality of expansions and maintenance needed in the two holy mosques in Makkah and Medina[1] and to avoid any mistakes or disasters that may occur. One of the main components of the total number of pilgrims is the Number of Pilgrims coming from outside the kingdom (NPO).

The first objective of this paper is to use the modern Bayesian approach to implement the identification, diagnostic checking and forecasting phases of NPO data. The foundation of the proposed Bayesian analysis is to use the pure autoregressive processes, denoted by AR (P) for short, to model and forecast the NPO data. There are three main reasons to use pure AR (P) processes to analyze our data. First, most data arise in real applications can be well presented by such processes. Second, the likelihood function OF pure AR (P) processes is analytically tractable because the white noise is a linear function of the model parameters, hence, as it will be seen in section 3, one may develop the exact posterior mass function of the model order. Third, it became clear to the authors, after a preliminary examination of the data, as it will be seen in section 4, that the pure AR (P) processes are appropriate to model and forecast the NPO data. The second objective of this paper is to use the well- known Box-Jenkins methodology to do a complete time series analysis of the NPO data. The final objective of this paper is to compare the accuracy of the results achieved by the Bayesian and Box-Jenkins approaches. he literature on time series is vast and can be found in many other areas other than statistics. Most of the literature is non- Bayesian and the reader is referred to Box-Jenkins (1970), Priestely (1981), Bowerman and O' Connell (1987), Tong (1990), Harvey (1993), Wei (2005) and Liu (2009). It is a fact that the methodology of Box-Jenkins is the most popular and prevailing traditional methodology to model and forecast time series. However, the Box-Jenkins methodology has serious disadvantages and drawbacks. Their identification technique is highly nonobjective and requires a very careful examination for the raw data and very good skills. In order to implement the identification stage, the time series analyst should be knowledgeable, well experienced and highly trained. In addition, he should have a large amount of data in order to identify an adequate model, see Chatfield (2004).

On the other hand, the Bayesian analysis of time series is still being developed and most of the Bayesian contributions have been occurred within the last three decades. Zellner (1971) introduced the subject for special autoregressive and econometric models. Newbold (1973) made an important contribution by his analysis of ARMA type transfer function

[1] **Makkah and Medina are the two well-known holy cities in Saudi Arabia where pilgrims should perform circumambulation in Makkah holy mosque: and most of them visit Medina mosque.*

models. Newbold's results were based on a t- approximation for the posterior analysis, as did the latter work of Zellner and Reynolds (1978). During this period, Bayesian forecasting was advanced by Chow (1975), who found the moments of the joint predictive distribution of future observations. One of the most important contributions of time series analysis was done by Monahan (1981), who used a numerical integration to implement the identification, estimation and forecasting phases of low order ARMA processes. This was the first Bayesian attempt to perform a numerical comprehensive time series analysis and was very valuable contribution. Shaarawy and Broemeling (1984) and Broemeling and Shaarawy (1988) have developed Bayesian techniques for identification, estimation, diagnostic checking and forecasting phases based on a t-approximation to the posterior distribution of the coefficients. Their first study has been extended later by Chen (1992) to bilinear model. Recently, Shaarawy and Ali (2003) have initiated a direct Bayesian technique to identify the orders of seasonal autoregressive processes. Their approach has been extended to the case of moving average processes by Shaarawy et al. (2007). The multivariate version of their direct approach has been introduced by Shaarawy and Ali (2008).

The Bayesian approach has several advantages when compared with Box-Jenkins approach, most obvious is pedagogical. It is much easier to learn the Bayesian methodology once one has mastered the inferential interpretation of Bayes' theorem. On the other hand, with traditional analysis, one must learn a large variety of sampling theory techniques. Second, the importance of Bayesian methods in economics, finance, engineering, education and other fields has increased rapidly over the last two decades. Third, the Bayesian methodology provides the time series analyst, in all areas of applications, with a formal and unifying way to incorporate the prior information in the analysis before seeing the data and this may lead to exact small sample results, see Broemeling and Shaarawy (1988). Fourth, our proposed Bayesian methodology does not assume stationarity.

The remainder of this paper is organized as follows: Section II presents autoregressive processes and processes and their basic characteristics. A complete Bayesian analysis for NPO is developed in section III. Section IV is devoted to model and forecast the NPO using the traditional method developed by Box and Jenkins (1970). Section V is dedicated to evaluate the forecast performance of Bayesian and Box-Jenkins procedures and compare their numerical results. Finally, the paper is concluded in Section VI.

II. AUTOREGRESSIVE PROCESSES

The autoregressive models are very useful in modeling time series data arise in many areas of scientific endeavor such as economics, business, marketing, physics, engineering and education.

Let $Y = [y(1) \;\; y(2) \ldots y(n)]'$ be a vector of n observations generated from autoregressive process of order p, denoted by AR(p) for short. The model has the form (see Box and Jenkins (1970))

$$\phi(B)y(t) = \varepsilon(t) \qquad (1)$$

Where B is the backshift operator defined by

$$B^r y(t) = y(t-r) \quad , r = 1, 2, \ldots$$

y(t) denotes the time series observations, t=1, 2, …, n, ε(t) denotes the random errors assumed to be i. i. d. $N(0,\tau^{-1})$, $\tau^{-} > 0$ is the precision parameter. Moreover

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

The AR(p) model is always invertible and is stationary if the roots of the polynomial equation $\phi(B) = 0$ lie outside the unit circle. The model (1) can be written in more explicit form as

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \ldots$$
$$+ \phi_p y(t-p) + \varepsilon(t)$$

$$y(t) = \sum_{i=1}^{p} \phi_i y(t-i) + \varepsilon(t) \quad ,$$
$$t = \ldots, -2, -1, 0, 1, 2, \ldots \qquad (2)$$

The vector $\phi(p) = (\phi_1 \quad \phi_2 \ldots \phi_p)'$ is the vector of the unknown coefficients. In practice the order p is unknown and one has to estimate it using the vector of observations $Y=[y(1) \quad y(2) \quad \ldots y(n)]'$. The Bayesian identification technique assumes that the order p is an additional parameter for which the marginal posterior probability mass function should be developed in a convenient form. The model (2.2) can be written in matrix notation as

$$y(p) = X(p)\phi(p) + \varepsilon(p) \qquad (3)$$

Where y(p) is a vector of order (n-p) with i-th element equal to y(p+i) and X(p) is a matrix of order (n-p)×p has the form

$X(p) = [y(t-1) \quad y(t-2) \; \ldots y(t-p)]'$ In addition, y(t-1) is a vector of order (n-p) with i-th element equal to y(p+i-1), y(t-2) is a vector of order (n-p) with i-th element equal to y(p+i-2), … and y(t-p) is a vector of order (n-p) with i-th element equal to y(i). The vector $\phi(p)$ is the vector of the unknown coefficients has the form $\phi(p) = [\phi_1(p) \quad \phi_2(p) \;\; .\phi_p(p)]'$ Finally, ε(p) is the vector the random errors of order (n-p) with i-th element equal to $\varepsilon(p+i)$. It is very important to mention that the vector y(p) and the matrix X(p) depend on the unknown order p. this means that for each value of p, say p_0, there is a corresponding vector $y(p_0)$ and a corresponding matrix $X(p_0)$.

III. BAYESIAN ANALYSIS OF NPO DATA

A. Identification

The time series of number of pilgrims coming from outside the kingdom of Saudi Arabia (NPO) (as shown in table (I)) consists of 44 observations (from year 1390AH[2] to year 1433 AH).

TABLE I. NUMBER OF PILGRIMS COMING FROM OUTSIDE THE KINGDOM (NPO)

Year	NPO	Year	NPO	NPO	Data	Year	NPO
1390	431270	1401	879368	1412	1012917	1423	1431012
1391	479339	1402	853555	1413	992813	1424	1419706
1392	645182	1403	1003911	1414	995611	1425	1534759
1393	607755	1404	919671	1415	1043274	1426	1557447
1394	918777	1405	851761	1416	1080465	1427	1654407
1395	894573	1406	856718	1417	1168591	1428	1707814
1396	719040	1407	960386	1418	1132344	1429	1729841
1397	739319	1408	762755	1419	1056730	1430	1613965
1398	830236	1409	774560	1420	1267555	1431	1799601
1399	862520	1410	828993	1421	1363992	1432	1828195
1400	812892	1411	720102	1422	1354184	1433	1752932

Let y(t) denotes the original series. We have found that the series y(t) is non-stationary and follows autoregressive scheme. The main objectives of this section is to identify the order of the series y(t), performing the diagnostic checking tests, and forecast the future observations using Bayesian approach. Given the assumptions outlined in the previous section, the conditional likelihood function of the process y(t) may be written as

$$L(\phi(p), p, \tau \mid y) \propto (\tau / 2\pi)^{[n-p]/2} \exp(-\frac{\tau}{2}[y(p) - X(p)\phi(p)]' \; [\; y(p) - X(p)\phi(p)]),\; \phi(p) \in R^{p},\; \tau > 0,\; p = 1, 2, ..., k \qquad (4)$$

Where k is the known maximum value of the order of the process and X(p) is the same as defined in the previous section. Shaarawy and Ali (2003) developed the Bayesian identification analysis using a Normal. Gamma prior for the parameters $\phi(p)$ and τ . Here, we assume that the conditional prior distribution of the parameters $\phi(p)$ and τ given p is Jeffreys' non-informative prior defined by

$$g(\phi(p), \tau \mid p) \propto \tau^{-1},\; \tau > 0 \qquad (5)$$

With respect to the prior probability mass function of the order p, we will assume that

$$\beta_i = P_r\,[\,p = i\,] \qquad ,i = 1,2,..., k \qquad (6)$$

The joint posterior distribution of the parameters $\phi(p)$, p and τ is proportional to the multiplication of the conditional likelihood function (4) and the priors (5) and (6). Integrating the joint posterior distribution of the parameters with respect to $\phi(B)$ and τ , one may prove that the marginal posterior probability mass function of the order p is

$$h\;(p \mid y) \propto \beta_i (\pi)^{-v(p)} \left| X(p) X'(p) \right|^{-1/2} \Gamma(v(p)) \{ y'(p) y(p) - y'(p) X(p) [X'(p) X(p)]^{-1} X'(p) y(p) \}^{-r(p)/2} \qquad (7)$$

Where

$$2r(p) = n - 2p \qquad ,p = 1, 2, ..., 4$$

The formula (7) has been used with the following three priors for the order p (assuming k=4)

Prior 1: $\beta_i = 1/4 \qquad ,i = 1,2,..., 4$

Prior2:

$$\beta_i \propto (0.5)^i \qquad ,i = 1,2,..., 4$$

Prior3:

$\beta_1 = 0.4$, $\beta_2 = 0.3$, $\beta_3 = 0.2$, $\beta_4 = 0.1$

The first prior assigns equal probabilities to the all possible values of the order p. The second prior is chosen in such a way to give probabilities that decline exponentially with the order, while the third prior is chosen in such a way to give probabilities that decrease with an amount 0.1 as the order increases. One may easily verify that the probabilities of the second prior are

$$\beta_1 = 0.5333 \;, \beta_2 = 0.2667$$
$$\beta_3 = 0.1333 \;, \beta_4 = 0.0667$$

The Matlab program has been used to do all computations required to calculate the marginal posterior probability mass function (7) for all three priors. The posterior probabilities are reported in table (II)

TABLE II. MARGINAL POSTERIOR PROBABILITY MASS FUNCTION OF THE ORDER FOR THE THREE PRIORS

Order	Jeffreys	Geometric	Arithmetic
1	1.0	1	1
2	0.0	0.0	0.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0

[2] *The years ate written using Islamic (Lunar) Calendar (AH).For instance, the year 1433 corresponds to year2012A.D. For more details see http://en.wikipedia.org/wiki/Islamic_Calendar*

Table (II) shows that the marginal posterior function attains its maximum at the first order with a perfect probability for all three priors. This means that the tentative adequate model is AR(1) for the series y(t) regardless the used prior.

IV. DIAGNOSTIC CHECKING

The next phase of Bayesian time series analysis is to check the model, which has been tentatively identified as AR (1), to see if it gives a reasonable fit to the data at hand. This has been accomplished by doing three different types of tests. The first type contains the test concerning the significance of the coefficient ϕ_1. The second type is to do the over fitting test. The third type is to do the residual analysis using the estimated residuals. With regard to the first type, the absolute value of the estimated parameter was 1.028. In addition, we have found that a 95% HPD interval for ϕ_1 was (0.995, 1.062). This means that we refuse the null hypothesis $H_0 : \phi_1 = 0$ and conclude that the parameter ϕ_1 is significant and the model AR (1) is appropriate for our data.

TABLE III. THE VALUES OF THE χ^2 STATISTICS

Lag(k)	Box-Pierce statistic	Ljung-Box statistic
12	10.088	12.603
24	18.084	28.155
36	19.798	34.625

Regarding the over fitting test, the higher model AR (2) has been fitted to the data. We know that the marginal posterior distribution of the parameters ϕ_1 and ϕ_2 is a non-central t with (n-4) degrees of freedom and location and precision parameters given by Broemeling and Shaarawy (1988). We have found that a 95% HPD interval for the added parameter ϕ_2 is (-0.027, 0.627) which conclude the zero value. Thus we cannot refuse the null hypothesis $H_0 : \phi_2 = 0$ and conclude that the identified model AR (2) is not appropriate for the data.

The third type diagnostic checking is to do residual analysis. If the fitted model AR(1) is appropriate, the calculated residuals $\hat{\varepsilon}(1), \hat{\varepsilon}(2), \ldots, \hat{\varepsilon}(n)$ should behave in manner which is consistent with the true model. This has been accomplished by doing several checks such as time series plot, the autocorrelation and partial autocorrelation functions of the residual, the portmanteau lack of fit test, and the autocorrelation function of the first difference of the residual. However, the time series plot of the residuals shows no outliers or any non-desirable autocorrelation or cyclic effect. The plot also gives no indication of a non-zero mean or non-constant variance. In addition, the autocorrelation and partial autocorrelation function of the residuals have no spikes. Moreover, The Anderson-Darling statistic for testing the normality assumption is 0.408 with p-value 0.333.These mean that the residuals resemble that of a whit noise sequence which supports the appropriateness of the identified model AR(1).

Instead of testing each autocorrelation, it is recommended to inspect the first k autocorrelation of the residual simultaneously using the Box-Pierce or Ljung-Box statistics. These two statistics have been calculated and the results are reported in table (III).

Comparing the results given by table (III) with the critical values of the χ^2 distribution with (k-1) degrees of freedom, we do not reject the null hypothesis

$$H_0 : \rho_\varepsilon(1) = \rho_\varepsilon(2) = \ldots = \rho_\varepsilon(k) = 0$$

for all values of k. These results support the appropriateness of the identified model AR(1). For more details about those two statistics, the reader is referred to Box and Jenkins (1970). Finally, the graph of the autocorrelation function of the first difference of the residuals cuts off after the first lag, while its partial autocorrelation function decays down. This means that the series of the first difference of the residuals has an MA(1) model with parameter does not differ significantly from1(see Box and Jenkins(1970)). This gives another support to the identified model AR(1).

B. *Forecasting*

The last phase of time series analysis is to forecast future observations. Thus, after passing through the modeling and diagnostic checking tests, confident that an AR(1) process has generated the NPO series, one would like to forecast Y(n+1), Y(n+2), …. The posterior predictive density of the future observations is the Bayesian tool to solve the forecasting problems.

a) One Step-Ahead Predictive Distribution

Assuming Jeffreys' non-informative prior (3.2) for the parameters ϕ_1 and τ , the predictive density of the next future observation y(n+1) can be shown to follow a non-central t distribution with (n-2) degrees of freedom, location

$$E[W(n+1) \mid y] = DE,$$

And precision

$$P[W(n+1) \mid Y] = D(n-2)(F - \frac{E}{D})^{-1}$$

Where

$$D = 1 - \frac{w^2(n)}{\sum_{t=1}^{n} w^2(t)} \quad ,$$

$$E = \frac{w(n)\sum_{t=2}^{n} w(t)w(t-1)}{\sum_{t=1}^{n} w^2(t)} \quad , \text{ and}$$

$$F = \sum_{t=2}^{n} w^2(t) - \frac{[\sum_{t=2}^{n} w(t)w(t-1)]^2}{\sum_{t=1}^{n} w^2(t)}$$

The posterior mean $E[Y(n+1)\mid y]$ provides a point forecast for the next observation y(n+1), and a $(1-\alpha)\%$ HPD interval for y(n+1) is

$$E[Y(n+1)\mid y)\pm t_{\alpha/2,n-2}P^{-1/2}[Y(n+1)\mid y]$$

b) Multi-Step-Ahead Predictive Distribution

The procedure followed throughout the above subsection to predict the first future observation y(n+1), using the one step-ahead predictive density, can be generalized. So that we can predict the k^{th} future observation y(n+k), using the k step-ahead predictive density. However, the prediction process of y(n+k) is conditional on the predictions of its preceding future observations y(n+1), y(n+2),.., y(n+k-1).

Thus, the forecasting process of the future observations should be employed step by step. One should first predict y(n+1) using the one step-ahead predictive density. Then, depending on a point forecast for y_{n+1} one can predict y(n+2) using the two step-ahead predictive density, which is conditional on the point forecast of y(n+1).

This process can be repeated for the succeeding future observations. For more details, see Broemeling and Shaarawy (1988).The model AR (1) has been used to forecast the next five future observations. The point forecasts and 95% HPD intervals for these observations are given by table (IV).

TABLE IV. THE FUTURE FIVE FORECASTS AND THEIR CONFIDENCE INTERVALS USING BAYESIAN PROCEDURE.

Year	Point forecast	HPD intervals
1434	1790544	(1571767, 2009320)
1435	1828962	(1612811, 2045113)
1436	1868204	(1654586, 2081823)
1437	1908289	(1697115, 2119464)
1438	1949234	(1740421, 2158047)

V. BOX AND JENKINS ANALYSIS OF NPO DATA

Box and Jenkins (1970) have presented a statistical analysis of ARMA(p,q) processes which has grown in popularity and is today the prevailing methodology of time series analysis. They assume that the time series at hand (or a transformation of the series) could be presented by a parsimonious stationary and invertible ARMA process such that one can perform the four phases of time series analysis: identification (order determination), estimation, diagnostic checking, and forecasting. In what follows we give a brief summary to each phase using Box and Jenkins methodology.

According to Box and Jenkins, the identification of the order p and q is done by computing the sample autocorrelation and partial autocorrelation functions and matching them with their theoretical counterparts, which are mathematically known for low-order processes.

Their methodology has been widely used and explained by many others such as Chatfield (1980), Priestely (1981), Bowerman and O'Connell (1987), Tong (1990), Harvey (1993), Wei (2005), Box et al.(2008) and Liu (2009).

After the model is tentatively identified, say an ARMA (p, q), the autoregressive parameters $\phi=(\phi_1 \quad \phi_2 \quad \dots \quad \phi_p)^T$, the moving average parameters $\theta=(\theta_1 \quad \theta_2 \quad \dots \quad \theta_q)^T$, and the residual variance σ^2 are estimated by maximum likelihood or nonlinear least squares methods. The maximum likelihood estimates (MLE) and least squares estimates (LSE) may be based on either the full (unconditional) likelihood function or a conditional likelihood function. These techniques are given in details by Priestely (1981). If q=0, the noise term is a linear function of the parameters and one may use the well-known linear least squares algorithm to estimate the parameters $\phi_1, \quad \phi_2, \dots \quad \phi_p$.

The third phase of a time series is to check the adequacy of the identified model to see if it gives a reasonable fit to the data at hand. This is mainly accomplished by a series of diagnostic checks using the estimated residuals. One may inspect the graphs of autocorrelation and partial autocorrelation to make sure that they do not have significant spikes particularly at low lags. In addition, one may investigate the residual plot to make sure that it does not have a particular pattern. Moreover, one may investigate the Ljung and Pierce statistic. One may also investigate the fitted model of the first differences of the residuals to see if it has the first order moving average model. For more details, see Box and Jenkins (1976) and Box et al. (2008).

The last phase of a time series analysis is to forecast future observations where the predicted observations are computed recursively from an estimated conditional expectation, namely, the conditional expectation of a future observation given the past data. For more details see, Box and Jenkins (1976) and Box et al. (2008).

The main objective of first section is to model and forecast the NPO data using the most popular well - known approach developed by Box and Jenkins in 1970. In order to achieve an adequate tentative model for the NPO data, the time series plot and the autocorrelation function (acf) are plotted in figures (I) and (II) respectively.

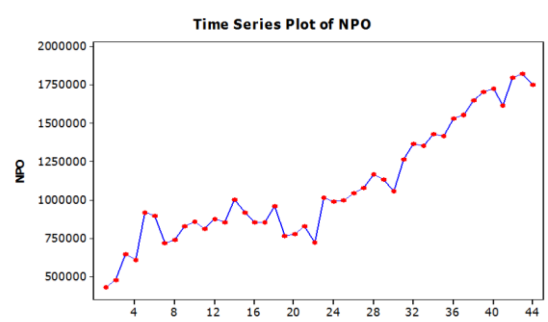

Fig. 1.

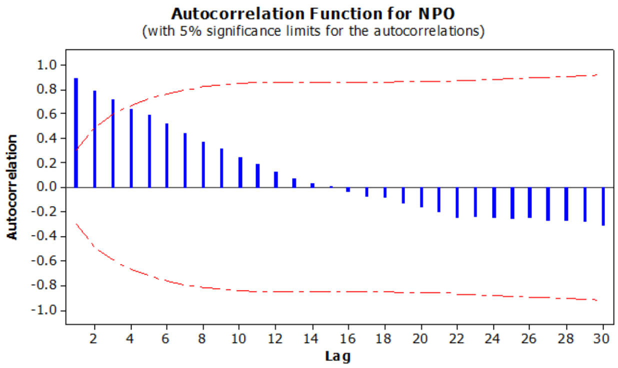

Fig. 2.

Inspecting the above graphs, it is easy to conclude that the NPO data y_t is non - stationary in the mean and variance.

In order to use Box and Jenkins methodology, it was necessary to use some mathematical transformation to convert the original data y_t into a new stationary series. As we have said before, this one of the disadvantages of Box and Jenkins methodology. After doing many trails, we can say that the second difference of the logarithm of the NPO data succeeded to convert the original time series y_t to a stationary one. Let z_t denotes to the new series, then

$$z_t = \log(y_t) - 2\log(y_{t-1}) + \log(y_{t-2}),\quad t = 3,4,\ldots,44$$

Figures (III) and (IV) show the time series plot and the autocorrelation function of the time series z_t.

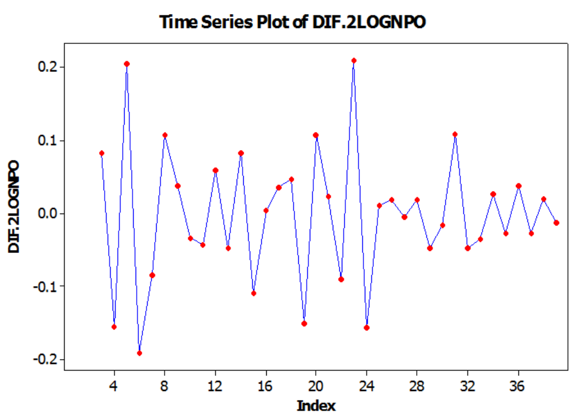

Fig. 3.

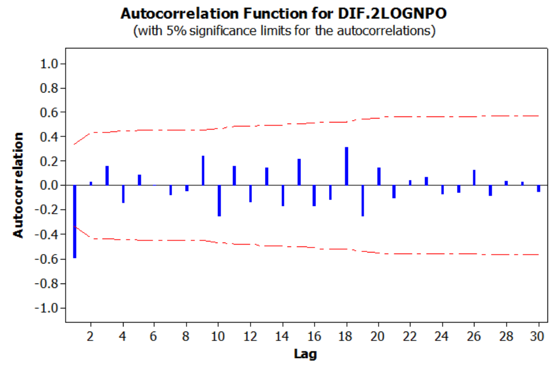

Fig. 4.

Inspecting these graphs, one may say that the time series z_t tends to be stationary. Thus, one may use Box and Jenkins to model and forecast the series z_t. In order to identify a tentative model to z_t data, the partial autocorrelation function is computed and its graph is given by figure (V).

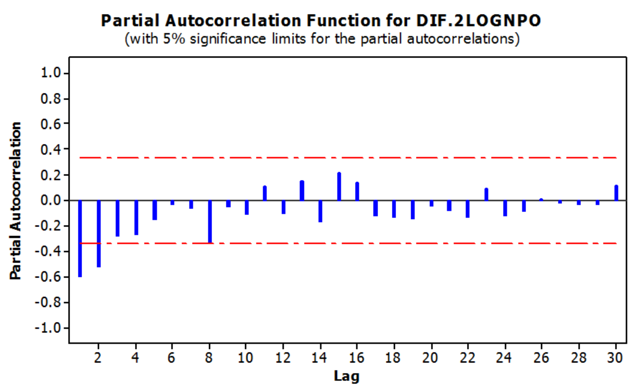

Fig. 5.

Inspecting the autocorrelation of the series z_t, one may notice that the coefficients of the autocorrelation function are small after the first lag. In addition, one may notice that the partial autocorrelation coefficients are small at the third and fourth lags and very small after the fourth lag. This means that we have four different models, one of them must be chosen in order to have good forecasts. The first choice is the first order moving average, denoted by MA(1), model , while the other three choices are AR(2), AR (3), and AR (4) models. We have started to analyze the series z_t using MA(1) model, but the numerical results were unsatisfactory because the model did not pass most of the diagnostic checking tests.

For example the autocorrelation function seems to have a spike at the first lag and the p-value for Anderson-Darling statistic was 0.014. Second, the AR(2) model was used to fit the data, but the numerical results of the diagnostic checking have shown that more autoregressive parameters should be added to the model. Therefore, we have used AR(3) model to fit the data, but it turned out that the model still needs more autoregressive parameters. Finally AR (4) model has been used to analyze the data, and all numerical results of the diagnostic tests were very satisfactory. Thus, we concluded that AR (4) model is the most adequate one the model and forecast the series Z_t , i.e. the ARIMA (4, 2, 0) is the most adequate model to fit the logarithm of the NPO data.

Finally, the model ARIMA(4,2,0) has been used to forecast the next five future observations. The point forecasts and 95% HPD intervals for these observations are given by table (V).

TABLE V. THE FUTURE FIVE FORECASTS AND THEIR CONFIDENCE INTERVALS USING BOX AND JENKINS PROCEDURE

Year	Point forecast	Confidence intervals
1434	1827212	(1423122 , 2346042)
1435	1828925	(1316379 , 2541038)
1436	1880008	(1247717 , 2832718)
1437	1896373	(1109448 , 3241462)
1438	1893155	(979379 , 3659497)

VI. A COMPARATIVE STUDY

This section has three main objectives. The first is to evaluate the forecast performance of Bayesian procedure outlined in section 3. The second objective is to evaluate the forecast performance of the Box-Jenkins procedure used in section 4.

The final objective is to compare the numerical results achieved by the two proposed approaches. In order to achieve the main goals, a small portion of the NPO data at the end of the data are reserved solely for forecast comparison. In statistical literature, these data are referred to as *hold-out sample, or post-sample*, and in principle are not used in model identification or estimation when evaluating forecast performance. However, there are several criteria to evaluate forecast performance of a model, including mean absolute percentage error (MAPE), mean absolute deviation (MAD), and root mean squared error (RMSE) as defined below

$$MAPE = \tfrac{1}{m}\sum_{t=1}^{m}\left|\frac{real\ value - forecast}{real\ value}\right|.100$$

$$MAD = \tfrac{1}{m}\sum_{t=1}^{n}\left|true\ value - forecast\right|$$

$$RMSE = \left[\tfrac{1}{m}\sum_{t=1}^{n}\left(true\ value - forecast\right)^2\right]^{1/2}$$

Where, m is the total number of observations in the *hold-out sample (post-sample).*

In order to use the above criteria to evaluate the forecast performance of the two proposed approaches and compare between them, the last 5 observations (about 12% of the whole data) are reserved as the *hold-out sample (post-sample).* The first 39 observations were used to forecast the next five observations using Bayesian and Box-Jenkins approaches; then the five forecasts were compared with the five real observations and the three above criteria were calculated. The results are reported in tables (VI) and (VII) respectively.

TABLE VI. THE FORECAST PERFORMANCE OF BAYESIAN APPROACH

Year	Real obs-ervations	Forecasts	APE%	AD	SE
1429	1707814	1756113	1.52	48299	2332793401
1430	1729841	1805778	11.88	75937	5766427969
1431	1613965	1856848	3.18	242883	58992151689
1432	1799601	1909363	4.44	109762	12047696644
1433	1828195	1963362	12.00	135167	18270117889
Mean			5.7000	122409.6	19481837518.4

TABLE VII. THE FORECAST PERFORMANCE OF BOX-JENKINS APPROACH

Year	Real observations	Forecasts	APE%	AD	SE
1429	1707814	1748546	1.08	40732	1659095824
1430	1729841	1852292	14.77	122451	14994247401
1431	1613965	1904772	5.84.	290807	84568711249
1432	1799601	1985503	8.61	185902	34559553604
1433	1828195	2066444	17.89	238249	56762586001
Mean			9.600	175628.2	38508838815.8

Inspecting the results given by table (I), one may conclude that the identified model AR(1) gives very good Bayesian forecasts since the mean absolute percentage error (MAPE) is very low, being equal to 5.7%. On the other hand, the corresponding value of MAPE computed by Box-Jenkins procedure is 9.6%. In addition the MAD and the RMSE for Bayesian approach were 122409.6 and 139577.4 respectively, while the corresponding values for Box-Jenkins approach were 175628.2 and 196236.7.

In addition, 95% confidence intervals intervals have been computed for the *hold-out sample (post-sample)* using the two proposed procedures and the results are reported in table (VIII) and (IX).

TABLE VIII. 95% CONFIDENCE INTERVALS OF THE LAST FIVE OBSERVATIONS USING BAYESIAN ANALYSIS

Year	Lower Bound	Upper Bound	Length of the Intervals
1429	1655613	2094008	438395
1430	1706121	2139107	432986
1431	1757749	2185525	427776
1432	1810534	2233287	422753
1433	1864513	2282418	417906
Mean			427962

TABLE IX. CONFIDENCE INTERVALS OF THE LAST FIVE FUTURE OBSERVATIONS USING BOX AND JENKINS PROCEDURE

Year	Lower Bound	Upper Bound	Length of the Intervals
1429	1423122	2346042	922920
1430	1316379	2541038	1224659
1431	1247717	2832718	1585001
1432	1109448	3241462	2132014
1433	97379	3659497	2680119
Mean			1708943

Comparing the numerical results of Bayesian approach, in forecasting the last five observations, with the numerical results of Box and Jenkins approach, one may conclude the flowing:

1) The mean absolute percentage error (MAPE) achieved by Box and Jenkins approach is higher than the corresponding value achieved by Bayesian approach with more than 68%.

2) The mean absolute deviation (MAD) achieved by Box and Jenkins approach is higher than the corresponding value achieved by Bayesian approach with more than 43%.

3) The root mean squared error (RMSE) achieved by Box and Jenkins approach is higher than the corresponding value achieved by Bayesian approach with more than 40%.

4) The 95% confidence interval for the next step ahead forecast achieved by Box and Jenkins approach is wider than the corresponding value achieved by Bayesian approach with more than 110%.

5) The mean of lengths of the 95% confidence intervals for the last five observations achieved by Box and Jenkins approach is higher than the corresponding value achieved by Bayesian approach with more than 299%.

From the foregoing numerical results, we conclude that the Bayesian approach is much more accurate than Box and Jenkins approach in modeling and forecasting the NPO data because it gives better forecasts and narrower confidence intervals.

VII. SUMMERY AND CONCLUSION

The authors have proposed to use the Bayesian approach to develop a complete time series analysis of number of Pilgrims coming from outside the Kingdom of Saudi Arabia from year 1390AH to year 1433AH. Using a Jeffreys' non-informative prior for the parameters and three different priors for the model order, the proposed methodology is to develop the marginal posterior probability mass function of the model order is given in an easy and convenient form using the approach developed by Shaarawy and Ali(2003) . Then, one may investigate the behavior of the marginal posterior probability mass function and choose the order at which the marginal probability mass function attains its maximum to be the identified order. We have found that AR(1) model is the identified tentative model for the series. The tentative model has passed all the diagnostic checking tests with high precision. Point forecasts and HPD intervals for the next five future years are provided by the authors using the marginal and conditional predictive densities given by Broemeling and Shaarawy (1988). In addition, the traditional Box and Jenkins approach was used to analyze the same data. The numerical results achieved by Bayesian approach were much better than the results achieved by the traditional Box and Jenkins approach.

REFERENCES

[1] Bowerman, B. and O'Connell (1987). Time Series Forecasting: Unified Concepts and Computer Implementation. Boston: PWS Publishers Dexbury Press.

[2] Box, G. and Jenkins, G. (1970). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco

[3] Broemling, L. and Shaarawy, S. (1988). Time Series Analysis: A Bayesian Analysis in the Time Domain. Bayesian Analysis of Time Series and Dynamic Model, edited by Spall, J.

[4] Chatfield, C. (1980). The Analysis of Time Series: Theory and Practice. Chapman and Hill Ltd, London.

[5] Chatfield, C. (2004). The Analysis of Time Series: An Introduction. 6th ed. Boca Raton, Fla. : Chapman & Hall/CRC, - Texts in statistical science.

[6] Chow, G. (1975). Multi-period Prediction from Stochastic Difference Equations by Bayesian Methods. Chapter 8 of Studies in Bayesian Econometrics and Statistics, edited by S.E. Fienberg and A. Zellner. North-Holland, Amsterdam.

[7] Harvey, A. (1993). Time Series Models, 2nd edition. The MIT Press.

[8] Liu, L. (2009), Time Series Analysis and Forecasting. 2nd edition. Scientific Computing Association Corp, USA.

[9] Monahan, J. (1983). Fully Bayesian Analysis of ARIMA Time Series Models, Journal of Econometrics, Vol. 21, pp. 307-331.

[10] Newbold, P.(1973). Bayesian Estimation of Box and Jenkins Transfer Function Model for Noise Models. Journal of the Royal Statistical Society, Series B, vol. 35. No. 2, pp. 323-336.

[11] Priestley, M. (1981). Spectral Analysis of Time Series, Academic Press, London.

[12] Shaarawy, S. and Ali, S. (2003). Bayesian Identification of Seasonal Autoregressive Models. Communications in Statistics-Theory and Methods, Vol. 32, Issue 5, pp.1067-1084.

[13] Shaarawy, S. and Ali, S. (2008). Bayesian Identification of Multivariate Autoregressive Processes. Communications in Statistics-Theory and Methods, Vol. 37, Issue 5, pp.791-802.

[14] Shaarawy, S. and Broemeling, L. (1984). Bayesian Inference and Forecasts with Moving Average Processes. Comm. In Statis. Vol.13, No. 15.

[15] Shaarawy, S., Soliman, E. and Ali, S. (2007). Bayesian Identification of Moving Average Models, Communications in Statistics-Theory and Methods, Vol. 36, Issue 12, pp. 2301-2312.

[16] Tong, H. (1990). Non-Linear Time Series: A Dynamical System Approach. New York: Oxford University Press.

[17] Wei, W.W.S. (2005). Time Series Analysis: Univariate and Multivariate Methods. Addison Wesley, Reading, MA.

[18] Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons. Inc, New York.

[19] Zellner, A. and Reynolds, R. (1978). Bayesian Analysis of ARMA Models. Presented at the Sixteenth Seminar on Bayesian Inference in Econometrics, June 2-3.