Prediction of life expectancy in Saudi Arabia by 2030 using ARIMA models

A Alshehri¹, ², F El-Halawany¹ and H Abu-Zinadah¹

¹University of Jeddah College of Science, Hamzah Ibn Al Qasim Street, Al Sharafeyah, Jeddah 23218, Kingdom of Saudi Arabia
²Corresponding author’s e-mail: aalshehri1022.stu@uj.edu.sa

Abstract. Life expectancy at birth (LEB) is a major factor for decision-making bodies when developing new healthcare policies or improving existing ones. This paper, with the help of R language, processes and examine the LEB data in Saudi Arabia from 1960 to 2012 using time-series analysis. To test the validity of the model, LEB data from 2013 to 2018 are used. The performance of the selected auto regressive integrated moving average (ARIMA) model has been calculated using several comparing criteria, commonly used in forecasting statistical evaluation, for fitness and prediction phases. Finally, the Saudi Arabian LEB data from 2018 to 2030 are forecasted and analyzed to correspond with the Saudi Vision 2030 framework.

1. Introduction
Life expectancy at birth (LEB) is defined as “a standardized summary measure, sometimes used as an overall gauge of health, based on a population’s age structure and mortality experience” [1]. Therefore, it is a population mortality index that theoretically describes the average number of years in which an infant will live, considering different mortality rates within each age group. Life expectancy has increased significantly over the last few decades in different parts of the world. While the average age in 1960 reached 52.5 years worldwide, the current average is around 72 years.

Recent research showed the importance of life expectancy predictions in long-term development plans to understand potential health trajectories in the future [2]. From the literature, several life expectancy prediction methods have been proposed [3], [4], [5]. Extrapolation is the most common approach in demographic forecasting. Extrapolative methods are essentially atheoretical; the only assumption is that the future will be a continuation of the past. The most commonly used extrapolation method is the univariate ARIMA model forecasting [6], [7], [8], [9].

In Saudi Arabia, life expectancy has increased from 51.69 years in 1969 to 75 years in 2018, an average annual growth of 0.76, which indicates improved prevention and medical care services in the kingdom. In 2016 the kingdom established a strategic framework to reduce Saudi Arabia's dependence on oil, diversify its economy, and develop public service sectors such as health, education, infrastructure, recreation, and tourism. One of the Saudi Vision 2030 goals in health is to increase Saudi individuals’ life expectancy from 74 (75 currently) to 80 years by 2030 [10].

Research studies of this caliber are not abundant. However, a study used demographic literature to discuss the feasibility of the life expectancy vision target and the effect of cardiovascular diseases, smoking, obesity, lifestyle, and related policies on life expectancy in Saudi Arabia. The study also analyzed countries whose life expectancies are in the 70s and estimated annual gains of less than 0.31
after reaching 70 years, thus is lower than Saudi Arabia’s target of reaching 80 years of life expectancy in 2030, concluding that an average yearly gain of 0.25 is achievable for Saudi Arabia, given its fundamental life expectancy drivers [11].

This research study aims to predict life expectancy in Saudi Arabia by 2030 using ARIMA models. In addition, the Saudi Vision 2030 objective of increasing life expectancy needs an extensive examination using high-quality datasets. A reliable study will enable decision-makers to implement an effective development plan to sustain an increased life expectancy and overall health development and achieve this goal.

2. Data
This study used data extracted from the World Health Organization (WHO) and published in the World Development Indicators (WDI) because of challenges in collecting sufficient data. The data obtained from the WDI are from 1960 to 2017, and those for the year 2018 are from the Saudi Ministry of Health (MOH) statistical yearbook (see Table 1) [12]. The ARIMA model requires a minimum of 50 observations [13]; the dataset used in this study has 59 observations. Therefore, the data is adequate for analysis. LEB data from 1960 to 2012 were selected for analysis, and data from 2013-2018 were retained for verification. Table 1 shows the LEB in Saudi Arabia from 1960 to 2018.

| Year | LEB | Year | LEB | Year | LEB | Year | LEB |
|------|-----|------|-----|------|-----|------|-----|
| 1960 | 45.638 | 1975 | 58.148 | 1990 | 69.078 | 2005 | 73.256 |
| 1961 | 46.145 | 1976 | 59.179 | 1991 | 69.522 | 2006 | 73.355 |
| 1962 | 46.665 | 1977 | 60.169 | 1992 | 69.945 | 2007 | 73.467 |
| 1963 | 47.207 | 1978 | 61.116 | 1993 | 70.349 | 2008 | 73.598 |
| 1964 | 47.781 | 1979 | 62.016 | 1994 | 70.733 | 2009 | 73.75 |
| 1965 | 48.408 | 1980 | 62.863 | 1995 | 71.099 | 2010 | 73.917 |
| 1966 | 49.107 | 1981 | 63.66  | 1996 | 71.445 | 2011 | 74.089 |
| 1967 | 49.888 | 1982 | 64.411 | 1997 | 71.766 | 2012 | 74.254 |
| 1968 | 50.751 | 1983 | 65.122 | 1998 | 72.06  | 2013 | 74.402 |
| 1969 | 51.692 | 1984 | 65.795 | 1999 | 72.327 | 2014 | 74.533 |
| 1970 | 52.703 | 1985 | 66.43  | 2000 | 72.561 | 2015 | 74.651 |
| 1971 | 53.771 | 1986 | 67.028 | 2001 | 72.759 | 2016 | 74.761 |
| 1972 | 54.872 | 1987 | 67.589 | 2002 | 72.92  | 2017 | 74.874 |
| 1973 | 55.981 | 1988 | 68.116 | 2003 | 73.051 | 2018 | 75 |
| 1974 | 57.078 | 1989 | 68.611 | 2004 | 73.159 |

3. Theoretical framework
A strategic planning framework, such as Vision 2030, is a structured process to provide long-term development plans. When planning such strategies, governments follow the decision-making support system. The first step is to evaluate the current situation using estimation methods and forecasting models. The most popular forecasting method is the Box–Jenkins technique, widely known as the ARIMA models, which follows a three-stage approach aimed at selecting an appropriate model to estimate and forecast a univariate time series. It has a well-established foundation in the theory of statistics and dynamic systems, with reliable predictions when the model is selected correctly.
3.1. ARIMA models
ARIMA models fall under the Box–Jenkins approach, and the terms “ARIMA models” and “Box–Jenkins models” can be used interchangeably. A nonseasonal ARIMA model includes three variables \((p, d, q)\) where \(p\) is the number of autoregressive terms, \(q\) is the number of lagged forecast errors in the prediction equation or the order of the moving-average model, and \(d\) is the number of differences needed for stationarity. A nonseasonal ARIMA model is given in equation (1):

\[
y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_q
\]

where \(y_t\) is the series;
\(y_{t-1}\) denotes the values of previous years;
\(\epsilon_q\) refers to the previous prediction errors.

The construction of an ARIMA model consists of three main steps. Model identification establishes the degree of differencing required to make the series stationary and determine \(p\) and \(q\). According to Wei [14], the autocorrelation function (ACF) and the partial autocorrelation function (PACF) are vital tools in identifying \(p\) and \(q\). The ACF is the linear correlation between \(y_t\) and \(y_{t-k}\) and represents the \(q\) term, which is the moving average, while the PACF is known as the partial linear correlation function between \(y_t\) and \(y_{t-k}, y_{t-1}, \ldots, y_{t-k+1}\) and represents the autoregressive term \(p\). The next phase estimates the model parameters using the maximum likelihood estimation process outlined in the Box–Jenkins model [15]. Finally, diagnostic checks are performed to determine the efficacy of the model and examine if residuals are identified as a white-noise process. As seen in Figure 1, the Box–Jenkins methodology is an iterative process to obtain the optimal model in forecasting future values.

![Figure 1. The Box–Jenkins (ARIMA) methodology flowchart.](image)

3.2. Stationarity
According to Milles [16], “the autoregressive-moving average (ARMA) class of models relies on the assumption that the underlying process is weakly stationary, which restricts the mean and variance to be constant and requires the autocovariances to depend only on the time lag.” A nonstationary time series can achieve stationarity using mathematical transformations, such as differencing, and can also
be determined using unit root tests. The stationarity test used in this paper is the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test.

3.2.1. Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The KPSS test determines whether a time series is stationary or nonstationary due to the existence of a unit root. It proposes the following hypotheses:

\[ H_0 : \text{Data is stationary.} \quad H_1 : \text{Data is not stationary.} \]

The KPSS test calculation is based on linear regression by splitting the series into three main components. The KPSS regression equation is shown in equation (2).

\[ y_t = \beta_t + \beta + \epsilon_1 \quad (2) \]

where \( \beta_t \) is the deterministic trend;
\( \beta \) is the random walk;
\( \epsilon_1 \) is the stationary error.

3.2.2. Differencing. The Box–Jenkins methodology assumes a stationary time series. Nonstationary time series can transform into a weakly stationary series via differencing, which refers to the change between consecutive observations in the original series. Equation (3) shows the first order differencing calculations.

\[ z_t = y_t - y_{t-1} \quad (3) \]

Usually, after the first difference is taken, the data will appear stationary; if not, we might have to difference the data a second time. Practically, it is unnecessary to go beyond second order differences as given in equation (4):

\[ w_t = z_t - z_{t-1} = y_t - 2y_{t-1} + y_{t-2} \quad (4) \]

3.3. Goodness of fit
The goodness of fit for the ARIMA models is measured using the following criteria:

3.3.1. The Akaike information criterion (AIC). The AIC compares the quality of a set of statistical models [17]. The calculation is as follows:

\[ \text{AIC} = T \log \left( \frac{\text{SSE}}{T} \right) + 2(k + 2) \quad (5) \]

where \( T \) refers to the observations used for estimation;
\( SSE \) is the sum of the squares of residuals;
\( k \) is the number of model predictors.

However, the AIC might be incorrect when the sample size is small; in particular, when \( \binom{n}{k} < 40 \), the corrected Akaike information criterion (AICc), is required [18]. The formula of AICc is given in equation (6).

\[ \text{AICc} = \text{AIC} + \frac{2(k + 2)(k + 3)}{T - k - 3} \quad (6) \]

3.3.2. Ljung–Box test. The Ljung–Box test determines the absence of a serial autocorrelation in the model residuals up to a specified lag. The test statistic is in the following formula:
where $T$ is the number of observations; $r_z$ is the estimated autocorrelation of residuals; $k$ is the autocorrelation lag of the residuals.

The Ljung–Box test proposes the following hypotheses:

$H_0$ : The model does not show a lack of fit. $H_1$ : The model shows a lack of fit.

### 3.3.3. Performance indicators.

Forecast accuracy is measured using some performance indicators:

- Root mean square error (RMSE): The calculation formula of RMSE is:

$$RMSE = \left( \frac{1}{n} \sum_{t=1}^{n} (y_{pred_t} - y_{org_t})^2 \right)^{1/2}$$

(8)

where $y_{pred_t}$ represents the predicted value at time $t$; $y_{org_t}$ refers to the actual value at time $t$; $n$ is the total predicted time.

- Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_{pred_t} - y_{org_t}}{y_{org_t}} \right|$$

(9)

- Forecast accuracy (FA): Calculate the proximity between the forecast value and the real value. The closer the FA value is to 100%, the greater the accuracy of the forecast. The calculation formula is given in equation (10):

$$FA = \left( 1 - \left| \frac{y_{pred_t} - y_{org_t}}{y_{org_t}} \right| \right) \times 100\%$$

(10)

### 4. Empirical analysis

#### 4.1. Identification of the ARIMA model

The analysis was conducted using R language (4.0.2), released on June 22, 2020.

![Figure 2. LEB in Saudi Arabia from 1960 to 2012.](image-url)
Saudi Arabia’s LEB has increased over the years, an indication of the development of the country’s healthcare system. In Figure 2, LEB from 1960 to 2012 shows a visible upward trend, and the series is nonstationary. Accordingly, first and second order differences are applied to detrend the series. Also, to further determine stationarity, the KPSS test is performed, as shown in Table 2.

Table 2. KPSS test results of training series, the first-order differencing, and second-order differencing.

|                        | Training series | First-order differencing | Second-order differencing |
|------------------------|-----------------|--------------------------|--------------------------|
| Test statistic         | 1.46            | 1.22                     | 0.36                     |
| p-value                | 0.01            | 0.01                     | 0.09                     |

The KPSS test results show both the original series and the first order differencing nonstationary under a 0.05 significance level. However, the second order differences are significant, with a p-value of 0.09. Thus, we have good evidence to reject the null hypothesis since the p-value is greater than 0.05 and assume that the time series is stationary. Figure 3 shows the second-order difference time series with no apparent trend.

Figure 3. Time-series plot after the second-order difference of LEB in 1960–2012.

Figure 4(a) shows that the ACF plot decays and dies down in a damp sine-wave fashion while the PACF plot in (b) cuts off after the second lag. The parameter selection rules imply that the AR (2) model might be appropriate [15]. The AR (2) model is fitted to the differenced series, and the parameters are significant, which means that the fitting model for Saudi Arabia’s LEB data is the ARIMA (2,2,0) model.

Figure 4. ACF and PACF of the second difference for LEB.
4.2. Estimation

Based on the model identification in 4.1, ARIMA (2,2,0) was estimated against other models. First, we overfit the model by adding an additional AR parameter. That is, fit an ARIMA (3,2,0) to the training data. Also, underfitting the model by removing an AR parameter results in fitting an ARIMA (1,2,0) model. Lastly, we introduced a moving average parameter and fit an ARIMA (2,2,1) to evaluate whether the MA term produces better fitting results. To compare the quality of each model, the AIC and AICc were evaluated.

Table 3 shows that the minimum values for AIC and AICc are achieved when p = 3 and q = 0, which means ARIMA (3,2,0) is the most suitable model.

| ARIMA (p, d, q) | ARIMA (1,2,0) | ARIMA (2,2,0) | ARIMA (3,2,0) | ARIMA (2,2,1) |
|----------------|--------------|--------------|--------------|--------------|
| AIC            | −326.18      | −405.55      | −406.86      | −406.19      |
| AICc           | −325.93      | −405.04      | −405.99      | −405.32      |

The following are the results of the ARIMA (3,2,0) model fitting:

| ARIMA (3,2,0) | Coefficient | Std. Error | P-value |
|---------------|-------------|------------|---------|
| AR (1)        | 2.06011     | 0.13272    | 2.2E-16 |
| AR (2)        | −1.35626    | 0.25499    | 1.044E-07|
| AR (3)        | 0.2521      | 0.13566    | 0.06312 |

BIC = −399.13 log likelihood = 207.43

Table 4 shows the coefficients of the autoregressive terms. In R, no constant is permitted for d > 1, as a quadratic or higher-order pattern is particularly dangerous when forecasting [19]. The AR (1) and AR (2) coefficients are statistically significant and have predictive power. However, the AR (3) term is nonsignificant and is thus omitted from the prediction equation (see equation (11)).

\[
\hat{y}_t = 2.06 y_{t-1} - 1.36 y_{t-2} + \epsilon_t
\]  

4.3. Diagnostic check

After estimating the model, we will calculate the residuals of the fitted model to determine its performance and verify its reliability. According to Box–Jenkins [15], the diagnostic check involves testing the error terms’ statistical properties, the normality assumption, and weak white-noise assumption.
Figure 5. Residuals of the ARIMA (3,2,0) model.

Figure 5(a) shows the Q-Q plot of the residuals. The observations fall close throughout the line, indicating normally distributed errors. Also, from (b), the histogram and density plot have a normal distribution shape and a mean of zero. The residuals line plot in (c) fluctuates around zero with a consonant variance, suggesting that the residuals are white noise. In (d), the ACF lags are within the threshold limits, indicating no autocorrelation between the residual errors; generally, the residuals of the fitted model appear to be white noise as well. However, the plot indicates a substantial increase at lag one; thus, the Ljung–Box test is applied to check for serial autocorrelation and model fit. The results are shown in Table 5.

Table 5. Ljung–Box test of residuals.

| lags | Test statistic | P-value |
|------|----------------|---------|
| 7    | 7.00           | 0.43    |

The p-value of the Ljung–Box test is higher than the 0.05 level of significance, providing strong evidence against the null hypothesis. Therefore, the autocorrelations of the residuals are extremely small, and the first lag does not significantly affect model fitting, and the model does not show a significant lack of fit. As a result, ARIMA (3,2,0) is adequate to be used for further analysis. Diagnostic checking has shown that ARIMA (3,2,0) model met the assumptions. Thus, the model is used to forecast values for the next years. The out-of-sample from 2013 to 2018 forecasts are shown in Table 6.
Table 6. Actual LEB values and out-of-sample forecasts.

| Year | Actual | Forecasts | Std Error |
|------|--------|-----------|-----------|
| 2013 | 74.40  | 74.40     | 0.01      |
| 2014 | 74.53  | 74.52     | 0.04      |
| 2015 | 74.65  | 74.62     | 0.11      |
| 2016 | 74.76  | 74.68     | 0.22      |
| 2017 | 74.87  | 74.71     | 0.39      |
| 2018 | 75.00  | 74.71     | 0.63      |

Three performance indicators (RMSE, MAPE, and FA) are used to evaluate the model for further verification. The predicted effects of the model from 2013 to 2018 are shown in Table 7.

Table 7. Evaluation results.

| ARIMA (3,2,0) | RMSE   | MAPE   | FA      |
|---------------|--------|--------|---------|
| Test set      | 0.1431 | 0.1312 | 99.873% |

According to the evaluation results, ARIMA (3,2,0) has relatively small RMSE and MAPE values, and the FA value is close to 99.9%, indicating accurate model prediction results. Therefore, the ARIMA (3,2,0) model is credible in predicting future LEB values. A possible reason for the ARIMA model's improved accuracy is that their parameterization is done by minimizing the AIC and AICc criterion, which avoids over-fitting by considering both goodness of fit and model complexity [20].

Figure 6.

Figure 6 shows that the ARIMA (3,2,0) LEB forecast model proposed in this paper (red line) maintains a smooth error with the actual series in the first three years and accurately captures the direction of the future series. After the first three years, the forecasts drift steadily from the actual series values and appears to be biased downwards. However, the differences between the forecast and actual values are small.

4.4. Forecasting

The ARIMA (3,2,0) model is finalized after testing the model’s goodness of fit and verifying its accuracy.

The finalized model is used to predict LEB by 2030, with prediction intervals of 95% against all data (see Table 8).
Table 8. LEB forecasts with 95% prediction intervals.

| Year | Forecast | Std Error | 95% Prediction Intervals |
|------|----------|----------|--------------------------|
|      |          |          | Lo 95                   | Hi 95             |
| 2019 | 75.15    | 0.01     | 75.13                   | 75.17             |
| 2020 | 75.32    | 0.04     | 75.24                   | 75.40             |
| 2021 | 75.52    | 0.10     | 75.31                   | 75.72             |
| 2022 | 75.74    | 0.21     | 75.33                   | 76.16             |
| 2023 | 75.99    | 0.37     | 75.26                   | 76.73             |
| 2024 | 76.27    | 0.59     | 75.10                   | 77.43             |
| 2025 | 76.56    | 0.88     | 74.83                   | 78.28             |
| 2026 | 76.86    | 1.23     | 74.45                   | 79.27             |
| 2027 | 77.17    | 1.64     | 73.95                   | 80.40             |
| 2028 | 77.49    | 2.12     | 73.34                   | 81.64             |
| 2029 | 77.81    | 2.65     | 72.62                   | 83.00             |
| 2030 | 78.13    | 3.23     | 71.79                   | 84.46             |

The LEB forecasts generally increase with an average gain of 0.26 per year. The results are similar to the study conducted in 2018, which reported an annual gain in life expectancy by 0.25 [11]. In 2030, LEB is predicted to reach 78 years, with 95% prediction intervals from 71.79 to 84.46, rising by 3.13 years from 2018. In Figure 7, forecasts from 2019 to 2030 are plotted with 95% prediction intervals.

Figure 7. LEB series and predictions in 2019–2030.

5. Summary
Life expectancy has always been an important indicator of population health. Life expectancy forecasts for individuals in Saudi Arabia can support health program development in the kingdom and offer a scientific framework through which related policies can be assessed. Accurate predictions will enable decision-makers to implement effective development policies to sustain the overall rise in LEB. Accordingly, the 2030 Vision target of 80 years of life expectancy will be fulfilled.

Using the R language in this paper, we analyzed the life expectancy time series from 1960 to 2012 following the Box–Jenkins methodology. As a result, the ARIMA (3,2,0) model was constructed, and data from 2013 to 2018 were used for validation. The ARIMA (3,2,0) model succeeds in all diagnostic tests and has the lowest AIC and AIC0 values compared with other ARIMA models, with an MSE of
0.1312, an RMSE of 0.1431, and an FA of 99.9%. Consequently, the prediction performance of the model fitted in this paper is found to be sufficient and the forecasting results reliable.

Finally, we forecasted LEB in Saudi Arabia from 2019 to 2030, which continues to increase over the years and is expected to reach 78 years by 2030. Our findings are consistent with a recent study, which reported an annual rise of 0.25 in life expectancy in Saudi Arabia, reaching 78 years by 2030 [11].

Finally, this study is intended to offer a theoretical reference for assessing and adjusting related policies in the kingdom.

References
[1] Achey R L 2016 Life Expectancy. J. Palliat. Med. 19(4) 468
[2] Foreman K J, Marquez N, Dolgert A, Fukutaki K, Fullman N, McGaughey M, Pletcher M A, Smith A E, Tang K, Yuan C et al. 2018 Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. Lancet 392(10159) 2052-90
[3] Booth H 2006 Demographic forecasting: 1980 to 2005 in review. Int. J. Forecast. 22 547-581
[4] Shang H L 2012 Point and interval forecasts of age-specific life expectancies: A model averaging approach. Demogr. Res. 27 593-644
[5] Alho J M and Spencer B D 2005 Statistical demography and forecasting (New York: Springer) p 198-216
[6] Hussein E M and Mohamed E E 2015 Trend of Life Expectancy in Sudan Int. J. Sci. Res. 4(2) 2375-9
[7] Torri T and Vaupel J W 2012 Forecasting life expectancy in an international context. Int. J. Forecast. 28(2) 519-31
[8] Ntamjokouen A, Haberman S and Consigli G 2014 Modeling multi-population life expectancy: a cointegration approach. Insurance markets and companies: analyses and actuarial computations 5(2) 12-23
[9] Cao X, Hou Y, Zhang X, Xu C, Jia P, Sun X, Sun L, Gao Y, Yang H, Cui Z et al. 2020 A comparative, correlate analysis and projection of global and regional life expectancy, healthy life expectancy, and their GAP: 1995-2025. J. Glob. Health 10(2)
[10] KSA., Vision 2030. Ministry of Planning and Economic Development Riyadh 2016
[11] Bah S 2018 How feasible is the life expectancy target in the Saudi Arabia vision for 2030?. East. Mediterr. Health J. 24(4) 401-4
[12] Ministry of Health. Statistical book of 2019, Ministry of Health Riyadh 2019
[13] Shaarawi S M 2005 Introduction to Modern Time Series Analysis (Jeddah: King Abdulaziz University Press) p 357
[14] Wei W W 2006 Time Series Analysis: Univariate and Multivariate Methods (Boston: Pearson Addison Wesley) p 109
[15] Box G E, Jenkins G M, Reinsel G C and Ljung G M 2015 Time series analysis: forecasting and control (Hoboken: John Wiley & Sons) p 179-245
[16] Mills T C 2019 Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting (London: Academic Press) p 57
[17] Burnham K P and Anderson D R 1998 Practical Use of the Information-theoretic Approach In Model Selection and Inference (New York: Springer) p 48
[18] Yang X S 2019 Introduction to Algorithms for Data Mining and Machine Learning (London: Academic Press) p 87
[19] Hyndman R J and Khandakar Y 2008 Automatic time series forecasting: The forecast package for R. J. Stat. Softw. 27(1) 1–22
[20] Makridakis S, Spiliotis E and Assimakopoulos V 2018 Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one 13(3)