Trend of Life Expectancy in Sudan

Dr. Elfarazdag Mahjoub Mohammed Hussein, Dr. Elsiddig Edriss Mohamed

Department of Statistics-Faculty of Science-Tabuk University (KSA)

Abstract: ***The main objective of this paper is to construct a time series model to monitor and forecast the life expectancy at birth in Sudan. To achieve this objective, a series of life expectancy at birth ranged from 1990 to 2013was obtained from the Sudan Central Bureau of statistics and United Nation Annual Reports, time series analysis technique mainly box and Jenkins were used to find the required model .Analysis done by e-views package. The paper conclude that The most proper time series model to forecast the life expectancy at birth in Sudan is the ARIMA model, The highest forecasted Life Expectancy at Birth in the Sudan for the coming ten years will be attained in year 2023 as 66.4835 years and finally The life expectancy at Birth in Sudan is increasing at a decreasing rate.***

Keywords: Life expectancy, life expectancy index, box and Jenkins, auto regressive moving average Sudan

1. Introduction

Life expectancy (LE) is the expected number of years of life remaining at a given age (x), while the Life Expectancy Index (LEI) is a statistical measure used to determine the average lifespan of the population of a certain nation or area.

$$LEI = (Actual - Min)/(Max - Min)$$

Where, the goalpost of the life expectancy was determined by the UN as (85) Years for the maximum and (25) years for the minimum.

The industrial and agricultural revolution, as well as increased income levels led to improved nutrition and better access to drinking water and sanitation which increase the Europeans population's life expectancy in the 20th century and then spread to all countries that are today considered as developed,

The aim of the study is to find out a time series model based on annual basis for the life expectancy at birth in Sudan for the period 1990-2013. The importance of this study concentrated at the serious need of constructing a time series model for the future forecast to detect the pattern of change in the life expectancy at birth which helps to analyze the factors that affecting the Sudan life expectancy at birth.

2. Theoretical Framework

Time series analysis techniques, namely Box and Jenkins technique was used, the series extends over the period 1990-2013, which is fairly long; Autoregressive Integrated Moving Average (ARIMA) was chose forthe analysis.

2.1 Time Series Components

Time series consists of several components, which are:
1) Trend
2) Cyclical Variations.
3) Seasonal Variations.
4) Irregular fluctuations.

2.2 Time Series Decomposition Model

If a time series exhibits trend effects and seasonal effects, it can be useful to decompose it in order to isolate these effects. One model that allows us to do this is the multiplicative decomposition model, it's the most popular decomposition model, and it's expressed as follows:

$$Y_t = T_t * S_t * C_t * I_t$$

Also there is decomposition model known as the additive model, which expressed as follows:

$$Y_t = T_t + S_t + C_t + I_t$$

Where:
Y_t: The observed value of the time series in time period t.
T_t:The trend components in time period t.
S_t:The seasonal components in time period t.
C_t:The cyclical componentsin time period t.
I_t: the erratic components in time period t.

2.3 Forecasting Methods

There are many forecasting methods that can be divided into two basic types which are:

a) **Qualitative Forecasting Methods**
Qualitative forecasting methods generally use the opinion of the expert to subjectively predict future events.

b) **Quantitative Forecasting Methods**
Quantitative forecasting models are grouped into two main models, which are:

(I) Univariate Models
Univariate models predict the future events of time series on the basis of the past values of the time series (Powerman, 1979). When a univariate model is used; historical data are analyzed in an attempt to identify a data pattern, then assuming that it will continue in the future. Univariate forecasting models are most useful when conditions are expected to remain the same.

(II) Causal Models
The use of such models involves the identification of other variables that are related to the variable to be predicted, once these related variables have been identified, a statistical model that describes the relationships between these variables and the variable to be forecasted is developed. The statistical relationship derived is then used for forecasting the variable of interest.

Generally we can say that quantitative forecasting methods are used when historical data are available univariate models predict future values of the variable of interest on the basis of historical pattern of that variable, assuming the historical pattern will continue; causal models predict future values of the variable of interest based on the relation between that variable and other variables. Qualitative forecasting techniques are used when historical data are scarce or not available at all and depend on the opinions of experts

2.4 Choosing the Forecast Technique

In choosing the forecasting technique the forecaster must consider the following factors

1. The nature of the study variable.
2. The time frame.
3. The pattern of data.
4. The cost of forecasting.
5. The accuracy desired.
6. The availability of data.
7. The ease of operation and understanding.

The first factor to be considered in choosing a forecasting method is the form in which the forecast is desired i.e. determine whether the forecaster will use point or interval forecast. The second factor that can influence the choice of forecasting method is the time frame of the forecasting situation. Forecast are generated for point in time may be a number of days , weeks, months, quarters or years in the future. This length of time is called the time frame; the length of the time frame is usually categorized as follows:

1. Immediate less than month.
2. Short term more than three months to less than two years.
3. Long term two years or more.

The length of the time frame influence the choice of forecasting technique, typically a longer time frame makes accurate forecasting more difficult. The pattern of data must also be considered when choosing forecasting model. Thus, it is important to identify the existing data pattern. One of the most important factor that affect the choice of forecasting technique is the desired accuracy of the forecast, the availability of information and last the ease with which the forecasting method is operated and understood is important.

2.5 The Box-Jenkins Methodology

This methodology developed by G. E. P. Box and G. M. Jenkins, consists of four basic steps. The first step, called tentative identification step, involves tentatively identifying a model. Once a model has been identified, we estimate the model parameters in the second step that called the estimation step. The third step is called the diagnostic checking step, here we check the adequacy of the model, if the model proves to be inadequate, it must be modified. When a final model is determined, we use the model to forecast future time series values; this fourth step is called the forecasting step.

There are many Box-Jenkins models; these models can be grouped into the following three basic classes:

(A) Autoregressive models.
(B) Moving average models.
(C) Mixed autoregressive- moving average models.

Box-Jenkins models are often called ARIMA models [Autoregressive Integrated Moving Average]. The Univariate Box – Jenkins models have proven to provide accurate forecast in short term forecasting applications.

2.6 Stationary and Non-Stationary Time Series

The classical Box – Jenkins models describe stationary time series, thus in order to tentatively identify a Box – Jenkins models, we must first determine whether or not a time series under investigation is stationary, if it is not, we must transform it into a series of stationary time series values either by using the log or the reciprocal. A time series is said to be stationary if the statistical properties such as mean and variance of time series are constant through time , if we have observed n values $y_1, y_2, y_3, \ldots, y_n$, of a time series, we can use a plot of these values against time to help us to determine whether the time series is stationary or not. If the n values seem to fluctuate with constant variation around a constant mean, then it is reasonable to believe that the time series is stationary, if the n values do not fluctuate with constant variation, then it is reasonable to believe that the time series is no stationary. If we decided that the time series is not stationary we can transform it from non-stationary to stationary by taking the first differences of the non-stationary time series.

2.7Unit Root Test -Augmented Dickey-Fuller (DF) Test

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Unit root test is designed to test whether the time series data are stationary or not. Stationary time series data have the following characteristics:

1) Constant mean.
2) Finite variance.
3) The Correlogram diminishes as lag length increases.

The DF unit root test is based on the following regression forms:

1. Without Constant and Trend

$$\Delta Y_t = \delta Y_{t-1} + \mu_t$$

2. With Constant $\Delta Y_t = \alpha + \delta Y_{t-1} + \mu_t$
3. With Constant and Trend

$$\Delta Y_t = \alpha + \beta T + \delta Y_{t-1} + \mu_t$$

Where: α is constant, β and δ are the coefficients of the model.

2.8DF Unit Root Test Hypothesis

$$H_0 : \delta = 0$$
$$H_1 : \delta \neq 0$$

2.9. Decision Rule

If t*>ADF critical value, accept the null hypothesis. (Unit root exist).
Ift* >ADF critical value, reject the null hypothesis. (Unit root not exist).

2.10Test of Serial Correlation

The Correlogram test the serial autocorrelation under the following
Hypothesis:
If Prob. > 0.05 we accept H_0, which means serial correlation does not exist.
If Prob. < 0.05 we reject H0, which means serial correlation exist.

3. Empirical Analysis

3.1 Calculation of Life Expectancies

The starting point for calculating life expectancies is the age-specific death rates of the population members which defined as the number of death in a particular age group per 1000 population in the age group.

$$ASDR = \frac{Y_i}{P_i} \times 1000$$

Where:
ASDR: Age specific death rate.
Y_i: Number of deaths for age group i.
P_i:Number of population for age group i.
The age-specific death rates are calculated separately for separate groups of data which are believed to have different mortality rates (e.g. males and females, and perhaps smokers and non-smokers if data is available separately for those groups) and are then used to calculate a life table, from which one can calculate the probability of surviving to each age.

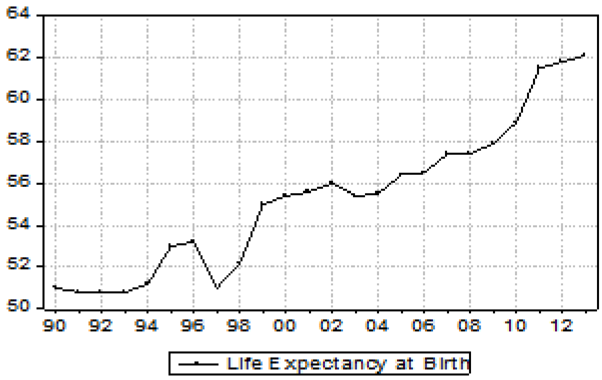

Figure 1: Life Expectancy at Birth
Figure (1) shows that the LE series is not stationary, there is an increasing trend.

Table 1: Augmented Dickey-Fuller Test

ADF Test Statistic	-6.900869	1% Critical Value*		-2.6889
		5% Critical Value		-1.9592
		10% Critical Value		-1.6246
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(LE,3)				
Method: Least Squares				
Sample(adjusted): 1994 2013				
Included observations: 20 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LE(-1),2)	-2.047083	0.296641	-6.900869	0.0000
D(LE(-1),3)	0.680290	0.194990	3.488846	0.0026
R-squared	0.766825	Mean dependent var		2.13E-15
Adjusted R-squared	0.753871	S.D. dependent var		2.347451
S.E. of regression	1.164604	Akaike info criterion		3.237279
Sum squared resid	24.41344	Schwarz criterion		3.336852
Log likelihood	-30.37279	F-statistic		59.19518
Durbin-Watson stat	2.247464	Prob(F-statistic)		0.000000

Table (1): Shows that the computed ADF test-statistics -6.900869is less than the critical values of "tau"(-2.6889, -1.9592, -1.6246)at 1%, 5% and 10% significant level respectively, the Durbin-Watson test is around 2, therefore we reject H_0 and accept H_1 which mean that the Life Expectancy at birth Series is stationary at the second difference .

Table 2: Second Difference Correlogram

Date: 02/19/15 Time: 23:37						
Sample: 1990 2013						
Included observations: 22						
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.**| . |	.**| . |	1	-0.218	-0.218	1.1984	0.274
****| . |	*****| . |	2	-0.518	-0.594	8.2865	0.016
. |* . |	.**| . |	3	0.172	-0.225	9.1039	0.028
. |* . |	.**| . |	4	0.154	-0.275	9.8008	0.044
.**| . |	***| . |	5	-0.203	-0.401	11.075	0.050
. |**. |	. | . |	6	0.225	0.052	12.739	0.047
. | . |	. *| . |	7	-0.022	-0.173	12.756	0.078
. *| . |	. | . |	8	-0.170	0.022	13.841	0.086
. |* . |	. | . |	9	0.088	0.049	14.155	0.117
. *| . |	***| . |	10	-0.080	-0.326	14.435	0.154
. | . |	. *| . |	11	-0.023	-0.152	14.461	0.209
. |**. |	. *| . |	12	0.254	-0.121	17.876	0.120

Table (2) shows that the Prob. Increases as the lag increase, which is a good indicator for the absence of serial autocorrelation at the second difference.

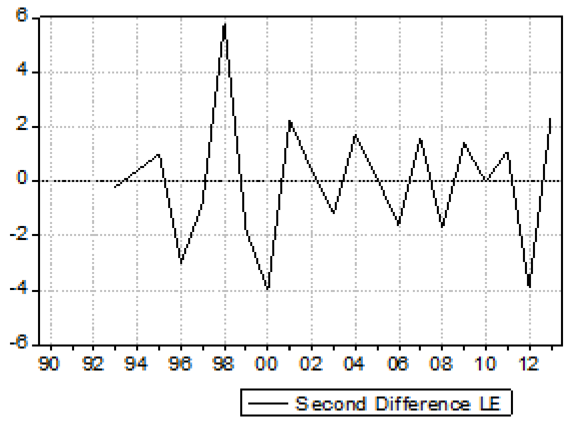

Figure 2: Second difference Life Expectancy at birth

Figure (2): shows that the Life Expectancy at birth series is stationary at the second difference.

Table 3: Model Estimation

Dependent Variable: LE				
Method: Least Squares				
Date: 02/20/15 Time: 00:06				
Sample(adjusted): 1991 2013				
Included observations: 23 after adjusting endpoints				
Convergence achieved after 11 iterations				
Backcast: 1989 1990				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	48.88873	0.925353	52.83250	0.0000
@TREND	0.533175	0.061732	8.636868	0.0000
AR(1)	0.638472	0.226874	2.814209	0.0111
MA(2)	-0.682086	0.250851	-2.719092	0.0136

International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

R-squared	0.953682	Mean dependent var	55.46957
Adjusted R-squared	0.946369	S.D. dependent var	3.526192
S.E. of regression	0.816607	Akaike info criterion	2.589453
Sum squared resid.	12.67009	Schwarz criterion	2.786931
Log likelihood	-25.77871	F-statistic	130.4040
Durbin-Watson stat	1.830922	Prob(F-statistic)	0.000000
Inverted AR Roots	.64		
Inverted MA Roots	.83	-.83	

Table (3) shows the estimated coefficients are statistically significant under a 5% level of significance. The overall regression fit, as measured by the R^2 statistics (R^2=0.953682), indicate a good fit. Since the Durbin Watson value is (1.830922) which is around **(2)** it means that there is no serial autocorrelation. The Akaike, Schwarz criteria (2.589453, 2.786931) indicate that the C @trend AR (1), MA (2) model should be preferred because they have the least values among the different models which can be fitted. The Prob. (**F-statistics**=0.000000) indicate that the whole model is statistically significant under 5% level of significance.

Estimation Command

LS LE C @TREND AR (1) MA (2)

Estimation Equation

LE = C(1) + C(2)*(@TREND) + [AR(1)=C(3),MA(2)=C(4),BACKCAST=1991]

Substituted Coefficients

LE = 48.88872796 + 0.5331749871*(@TREND) + [AR (1)=0.63847151,MA(2)= -0.6820864342,BACKCAST=1991]

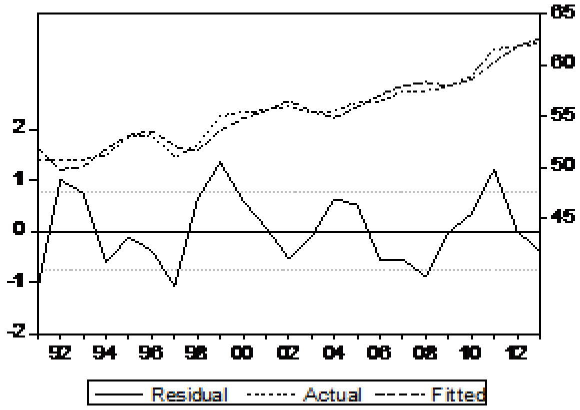

Figure 3: Fitted, Actual and Residual Life Expectancy at Birth

Figure (3) shows that the fitted values have no significant difference from the actual one.

Table 4:Actual,Fitted Life Expectancyat Birth

Year	Actual Life	Estimated Life
1990	51.00000	48.88873
1991	50.80000	49.42190
1992	50.80000	49.95508
1993	50.80000	50.48825
1994	51.20000	51.02143
1995	53.00000	51.55460
1996	53.20000	52.08778
1997	51.00000	52.62095
1998	52.20000	53.15413
1999	55.00000	53.68730
2000	55.40000	54.22048
2001	55.60000	54.75365
2002	56.00000	55.28683
2003	55.40000	55.82000
2004	55.50000	56.35318
2005	56.40000	56.88635
2006	56.50000	57.41953
2007	57.40000	57.95270
2008	57.40000	58.48588
2009	57.90000	59.01905
2010	58.90000	59.55223
2011	61.50000	60.08540
2012	61.80000	60.61858
2013	62.10000	61.15175

Table 5: Estimated Life Expectancy Index

Year	Estimated Life Expectancy	Estimated Life Expectancy Index	Change Rate of the Estimated Life Expectancy
1990	48.88873	0.398146	------
1991	49.42190	0.407032	0.022319
1992	49.95508	0.415918	0.021832
1993	50.48825	0.424804	0.021365
1994	51.02143	0.433691	0.020919
1995	51.55460	0.442577	0.02049
1996	52.08778	0.451463	0.020079
1997	52.62095	0.460349	0.019683
1998	53.15413	0.469236	0.019303
1999	53.68730	0.478122	0.018938
2000	54.22048	0.487008	0.018586
2001	54.75365	0.495894	0.018246
2002	55.28683	0.504781	0.01792
2003	55.82000	0.513667	0.017604
2004	56.35318	0.522553	0.0173
2005	56.88635	0.531439	0.017005
2006	57.41953	0.540326	0.016721
2007	57.95270	0.549212	0.016446
2008	58.48588	0.558098	0.01618
2009	59.01905	0.566984	0.015922
2010	59.55223	0.575871	0.015673
2011	60.08540	0.584757	0.015431
2012	60.61858	0.593643	0.015197
2013	61.15175	0.602529	0.014969

Table 6: Forecasted Life Expectancy at Birth and Forecasted Life Expectancy Index

Year	Forecasted Life Expectancy	Forecasted Life Expectancy Index	Change Rate of the Forecasted Life Expectancy
2014	61.68493	0.611416	0.014748
2015	62.21810	0.620302	0.014534
2016	62.75128	0.629188	0.014326
2017	63.28445	0.638074	0.014123
2018	63.81763	0.646961	0.013927
2019	64.35080	0.655847	0.013735
2020	64.88398	0.664733	0.013549
2021	65.41715	0.673619	0.013368
2022	65.95033	0.682506	0.013192
2023	66.48350	0.691392	0.01302

Tables (5 and 6): shows that the estimated and forecasted Life Expectancy at Birthwere increasing at a decreasing rate.

4. Conclusion

4.1 Results

1) Themost proper time series model to forecast the life expectancy at birth in Sudan is the ARIMA model.
2) The general trend of lifeexpectancy index in Sudan is increasing over time.
3) 3.The highest forecasted Life Expectancy at Birth in the Sudan for the coming ten years will be attained in year 2023 as 66.4835 years.
4) The life expectancy at Birth in Sudan is increasing at a decreasing rate.

4.2 Recommendations

Given the aforementioned findings, following policy recommendations are in order. Provision of free health services and improvement of primary health care at a higher rate than the one prevailed in the past by:

1. Not depriving the needy ones from getting appropriate health care.
2. Reducing maternal mortality rates, HIV and Malaria diseases.
3. Construction of new hospitals, health centres and offering training programmes for those who work in the medical fields such as doctors, nurses, laboratories technicians.

References

[1] Autocorrelation in Autoregressive-integrated moving Average Time Series Models", JASA, VOL.65, P. (1520-1526). (1970),
[2] Box, G.E.P and Jenkins, G.M., "Time Series Analysis Forecasting and control", Holden day, London. (1976)
[3] Box, G.E.P. & Pierce, D.A., "Distribution of the Residual
[4] Douglas, C .M & Contreas, J.G Note On Forecasting With Adaptive.
[5] Ferrer-i-Carbonell, A. and B.M.S. Van Praag, 2000. The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal. Presented at the Ninth European Workshop of Econometrics and Health Economics, September, Amsterdam, the Netherlands.
[6] http//www.developmentgoals.org.
[7] Kaiser, R. and Maravall, A. “Notes on Time Series Analysis ARIMA Models and Signal Extraction”, Benco de Espona servicio E studios(2001),
[8] Nuno Crato, “Some Results on the Spectral Analysis of stationary Time Series” Portugal Mathematic Vol. 53. Fasc. 2-1996.
[9] Sen, A,1999. A decade of Human Development. Key note speech. Human Development Report Office. UNDP.