ISSN 2709-4170 (Print)
ISSN 2791-0709 (online)
https://doi.org/10.52587/tmcs020102
Transactions in Mathematical and Computational Sciences
Vol. 02(1)(2022) pp. 01–09

TIME SERIES MODELING AND PREDICTION OF LIFE EXPECTANCY RATE AT BIRTH IN PAKISTAN

MUHAMMAD WASIM AMIR[1*], ALI HASSAN MALIK[2], ZEESHAN RAZA[3], MUHAMMAD SAAD KHAN[4]

[1]Govt. Associate College Miani, Sargodha, Pakistan
[2]Lahore Grammar School, Sargodha, Pakistan
[3]Govt. Elementary School No.1 Lalian, Chaniot, Pakistan
[4]The University of Lahore, Sargodha Campus, Sargodha, Pakistan

Abstract. The aim of this study attempted to shed light on the issues such as forecasting of life expectancy rate at birth in Pakistan. Data on life expectancy rate at birth are collected over 37 years from period 1980 to 2017. A variety of time series models are applied to find the most appropriate model for forecasting the life expectancy rate at birth (LEB) of Pakistan. In the study, the autoregressive integrated moving average ARIMA (3, 2, 2) is found to be the most adequate model for forecasting the LEB of Pakistan. The best model is selected based on the various model selection's tools. Furthermore, different test of runs and Jarque-Bera (JB) tests are used to justify the assumption of the randomness and normality of residuals. Based on the best ARIMA (3, 2, 2) model the forecasted value of the LEB in 2020, 2022, and 2025 are 67.05, 67.70, and 68.97 respectively.
Keywords: ARIMA, life expectancy, AIC.

1. Introduction

Life expectancy is a statistical measure of the average time a living thing is expected to live, based on the year of its birth, its current age, and other demographic factors including gender. The term life expectancy refers to the number of years a person can expect to live. Life expectancy is based on a computation of the average age that members of a specific population group will be when they die. Life expectancy refers to the average number

*Correspondence: muhammadwasim113@gmail.com
Received: 23.03.2021 Revised: 11.05.2022 Accepted: 14.06.2022 Published: 27.12.2022

of years a newborn is expected to live if mortality patterns at the time of its birth remain constant in the future. If a child is born in a country where the life expectancy is 75, he has a 75% chance to live. The countries with the highest value of life expectancy in the world are Hong Kong and China with a value of 84.28 and the country with the lowest value of life expectancy in the world is the Central African Republic, with a value of 51.38 (World Bank, [1]). The life expectancy at the time of birth in Pakistan is 66.48. The average value of the LEB in Pakistan is moderate because it is neither good nor bad. The LEB of Pakistan was greater than India and China in 1960 but in 2016 it changed; now it is less than the LEB of China and India. China has a life expectancy of 76.25 at birth and India has 68.56 (World Bank, [1]). The reason for low LEB includes unawareness, bad health facilities, and environmental conditions.

In the literature, various studies are conducted to check and forecast the future behavior of life expectancy rates, mortality rates and various others issues in different countries. Guillen and Vidiella-i-Anguera [2] used the Lee and Carter model to forecast the Spanish life expectancy and examined which factors are affecting forecasting. Olshansky *et al.* [3] conducted a study about the decline in life expectancy in the United States in the 21st century. Skiadas [4] applied the IM first exist time-series type model to forecast the females LEB in the Netherlands. Torri and Vaupel [5] forecasted the future situation of life expectancy by using the classic univariate ARIMA model. Bourg [6] showed that the life expectancy at birth will be increased in the upcoming decades. Bennett *et al.* [7] forecasted the age-specific mortality and life expectancy rate by using Bayesian spatiotemporal models. Shang [8] used the multilevel functional data method to forecast the mortality and life expectancy rate for a specific group of developed countries. Kontis [9] forecasted the life expectancy in 35 developed countries by using the Bayesian type method. Pascariu *et al.* [10] forecasted the life expectancy for men and women with the help of the Lee-Carter approach and the Cairns-Blake-Dowd strategy. Wu and Wang [11] forecasted the mortality rate by using the Gaussian regression model. Shair [12] used the functional data model to estimate the mortality and life expectancy at birth in developing countries. Nigri *et al.* [13] used four different time series models to forecast the life expectancy and lifespan disparity of the five countries. Perla *et al.* [14] applied the deep learning process to predict the future behavior of mortality rates.

Forecasting is indispensable for a country's future planning and policies to meet country requirements. The objective of the current study is to forecast the life expectancy rate at birth in Pakistan based on historical data from $1980-2017$ by using the appropriative time series model. The study also helps the government to build solid policies for Pakistan related to future life expectancy problems. The rest of the study is following as: Section 2 contains the structure of the ARIMA model along with their model selection criteria and diagnostics tests. Section 3 delivers the results of the different time series models and provides their complete discussion. The concluding remarks and recommendations are presented in section 4.

2. Methodology

The historical data for this study were collected from the World Bank. For forecasting purposes, different time series models are available in the literature and we are searching for the best one to forecast the future behavior of the LEB of Pakistan. A variety of time series models are applied to the historical data of the LEB and the Box-Jenkins type ARIMA model is considered the best for this data. The ARIMA model is symbolized by (p, d, q) where p is the order of the autoregressive process, d identifies the order of differencing, and q denoted the order of moving average. The mathematical structure of the ARIMA model is given below:

$$(1-\sum_{j=1}^{p}\phi_j L_j)(1-L)^d y_t = (1+\sum_{i=1}^{q}\theta_i L_i)e_t \tag{1}$$

The main issue in the ARIMA model is to identify the order of p,d,q. Generally, the order of the p identify by autocorrelation function (ACF) and order of q identify by partial autocorrelation function (PACF).

2.1. **Model specification.**

In time series forecasting the specification of the model is very important. The Akaike information criteria (AIC) and Schwarz Bayesian information criteria (SBIC) are commonly used to identify the finest model of time series among other counterparts. The mathematical function of the AIC is defined as:

$$AIC = nln(\hat{\sigma}_e^2) + 2M \tag{2}$$

where M is the number of parameters $M = p+q$ includes in the selected model and case of intercept include in the model the $M = p+q+1$. A model is considered best for further study if its AIC and SBIC values are minimum with other fitted models (Amir *et al.* [15]). The mathematical structure of SBIC mentioned by (Wei, [16]; Cooray, [17], Celik *et al.* [18]) is as:

$$SBIC = nln(\hat{\sigma}_e^2) + Mln(n) \tag{3}$$

2.2. **Time series model Diagnostics.**

After fitting the time series model the initial requirement is to justify the assumptions of the time series model such as normality, independence, and no autocorrelation of residuals of the best-fitted model (Amir *et al.* [19]). Different runs test, ACF, and PACF plots are applied to detect the autocorrelation within the residuals (Chatfield, [20]; Gujarati, [21]). The JB test is used to test the normality of the residuals.

2.3. **Forecasting accuracy measuring techniques.** The identification of the reliability of the forecast value is the key problem after selected the adequate model. To overcome this issue a variety of forecasting accuracy measure are available like Mean error (ME), mean percentage error (MPE), mean absolute PE (MAPE), mean absolute error (MAE), and root MSE (RMSE) that are used to checks the forecasting accuracy (Amir *et al.* [19]). The computation of these estimators is reported in Table 1.

TABLE 1. Forecast accuracy measure tools

Accuracy measuring tool	Formulation	References
MAE	$\text{MAE} = \frac{\sum_{t=1}^{n} \lvert u_t \rvert}{n}$	Amir *et al.* [19]
ME	$\text{ME} = \frac{\sum_{t=1}^{n} u_t}{n}$	Amir *et al.* [19]
MSE	$\text{MSE} = \frac{\sum_{t=1}^{n} u_t^2}{n}$	Amir *et al.* [19]
MPE	$\text{MPE} = \frac{\sum_{t=1}^{n} PU_t}{n}$	Amir *et al.* [19]
MAPE	$\text{MAPE} = \frac{\sum_{t=1}^{n} \lvert PU_t \rvert}{n}$	Amir *et al.* [19]

3. RESULTS AND DISCUSSION

In this study, a variety of time series models are applied to the LEB in Pakistan. The objective of applied these various models is to obtain reliable forecasts based on different statistical measures. Results of different time series models among their model selection and validity criteria are reported in Table 2. From Table 2, it is noted that the model $'M'$ shows the least value of AIC, and this model is considered the best model for forecasting the future behavior of life expectancy rate at birth among all other models based on historical data $1980-2012$. In Table 2, we also summarize the results of five different tests of runs on the residuals to determine whether the selected model is adequate for future prediction. It is interesting to note that the selected model passes all the tests and shows that the ARIMA $(3,2,2)$ is probably adequate for this data. The ARIMA $(3,2,2)$ model coefficient summary is provided in Table 3. (A)Random walk, (B) Random walk with drift $= 0.272313$, (C) Constant mean $= 61.529$, (D) Linear trend $= 57.0021 + 0.266289t$, (E) Quadratic trend $= 56.6885 + 0.320053t - 0.00158127t^2$, (F) Exponential trend $= exp(4.04486 + 0.00434056t)$, (G) S-curve trend $= exp(4.13747 + -0.151856/t)$, (H) Simple moving average of 2 terms, (I) Simple exponential smoothing with alpha $= 0.9999$, (J) Brown's linear exp. smoothing with alpha $= 0.9995$, (K) Holt's linear exp. smoothing with alpha $= 0.9999$ and beta $= 0.1349$, (L) Brown's quadratic exp. smoothing with alpha $= 0.9958$, (M) ARIMA$(3,2,2)$ (N) ARIMA$(2,2,3)$, (O) ARIMA$(3,2,3)$ Table 3 shows the result of the best fitted ARIMA $(3,2,2)$ model among their estimated co-efficient values. Based on Table 3, the estimated life expectancy forecasted model is as:

$$\nabla \hat{x}_t = 2.47068 \nabla \hat{x}_{t-1} - 2.26479 \nabla \hat{x}_{t-2} + 0.754789 \nabla \hat{x}_{t-3} + 2.47068 \nabla \hat{e}_{t-1} - 0.7 \nabla \hat{e}_{t-2} \quad (4)$$

TABLE 2. Life Expectancy at Birth forecasts (in percentage)

Model	RMSE	MAE	MAPE	ME	MPE	AIC	HQC	SBIC	RUNS	RUNM	AUTO	MEAN
(A)	0.27	0.27	0.44	0.27	0.44	−2.59	−2.59	−2.59	$***$	$***$		$***$
(B)	0.03	0.03	0.04	0.00	0.00	−6.81	−6.80	−6.77	$***$	$***$	$***$	$***$
(C)	2.58	2.19	3.58	0.00	−0.17	1.96	1.97	2.00	$***$	$***$	$***$	$***$
(D)	0.14	0.13	0.21	0.00	0.00	−3.75	−3.72	−3.66	$***$	$***$	$***$	OK
(E)	0.06	0.04	0.07	0.00	0.00	−5.49	−5.45	−5.36	$***$	$***$	$***$	OK
(F)	0.19	0.17	0.27	0.00	0.00	−3.21	−3.18	−3.11	$***$	$***$	$***$	OK
(G)	1.94	1.64	2.68	0.03	−0.05	1.44	1.47	1.53	$***$	$***$	$***$	$***$
(H)	0.42	0.41	0.66	0.41	0.66	−1.69	−1.68	−1.65	$***$	$***$	$***$	$***$
(I)	0.27	0.26	0.43	0.26	0.43	−2.53	−2.51	−2.48	$***$	$***$	OK	OK
(J)	0.08	0.03	0.04	0.02	0.03	−4.96	−4.94	−4.91	$***$	OK	OK	
(K)	0.02	0.02	0.03	−0.01	−0.01	−7.27	−7.24	−7.18	$***$	$***$	$***$	OK
(L)	0.08	0.02	0.04	0.00	0.00	−4.94	−4.92	−4.89	$***$	OK	OK	OK
(M)	0.00	0.00	0.00	0.00	0.00	−11.79	−11.72	−11.56	OK	OK	OK	OK
(N)	0.00	0.00	0.00	0.00	0.00	−11.72	−11.64	−11.49	OK	OK	OK	OK
(O)	0.00	0.00	0.00	0.00	0.00	−11.71	−11.62	−11.44	OK	OK	OK	OK

TABLE 3. Model Coefficient Summary of ARIMA $(3,2,2)$

Parameter	Estimate	stnd. Error	t-statistics	P-value
AR(1)	2.61377	0.132315	19.7542	0.00000
AR(2)	−2.64341	0.246364	−10.7297	0.00000
AR(3)	0.956952	0.157497	6.07601	0.00000
MA(1)	1.29961	0.0414719	31.3372	0.00000
MA(2)	−0.799518	0.0448894	−17.8108	0.00000

where δ shows the $2nd$ difference of the understudy variable, $\hat{x}_t$ is the forecasted LEB for time t years. To identify the autocorrelation between the residuals the ACF and PACF plots are presented in Figures 1 and 2. From Figures 1 and 2, it is observed that there is no autocorrelation within the fitted model residual series. The randomness of the residuals of the life expectancy at birth is checked using different three tests of runs. The results of the randomness tests are reported in Table 4. Another name of the randomness sequence is white noise. The first test identified the scenario of randomness by counting the number of times below or above the median. If the residuals are random then the number of such runs equals the expected number. The P-value of the 1st test greater than the margin of error of 0.05 which shows there is randomness in the residuals. The second test identifies the randomness of residuals by calculating the number of times runs rose or fell. The P-value of the $2nd$ test is also insignificant which discloses that the residuals of the ARIMA $(3,2,2)$ model are random. The 3rd test justifies the random sequences by using sum squares of the initial autocorrelation functions and the P-value concludes that the residual of the fitted model is random. We also accumulate the result of the JB test to justify the normality of

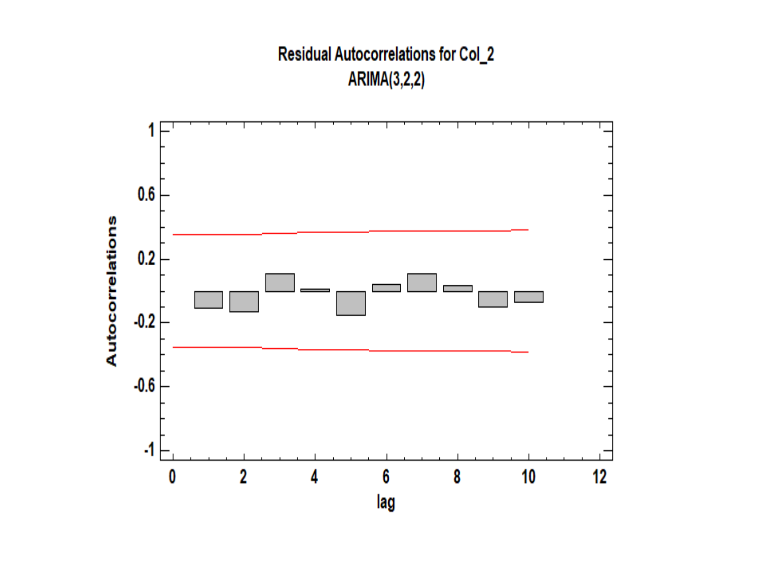

FIGURE 1. Residuals ACF Plot for ARIMA $(3,2,2)$

TABLE 4. Tests for residual normality and randomness

Test	Test Statistics	P-value	Expected number of runs
Runs overhead and underneath the median	0.185806	0.852592	16
Runs up and down	0.80483	0.420916	18
Box-Pierce Test	2.86383	0.852592	$--$
JB	$--$	0.503	$--$

fitted model residuals. From Table 4, it is noted that the P-value of the JB test is greater than the level of significance and the residuals of the ARIMA $(3,2,2)$ are normally distributed.

To check the efficiency of the fitted model one step ahead forecasts and their difference with actual observation are reported in Table 5. The forecasted values of the life expectancy rate at birth are presented in Table 6 and we noted that the life expectancy rate at the birth of Pakistan would become 68.97 in 2025. The predicted observations are based on the best-fitted statistical model and give the more reliable future behavior under the assumption that the environmental scenario will not be changed.

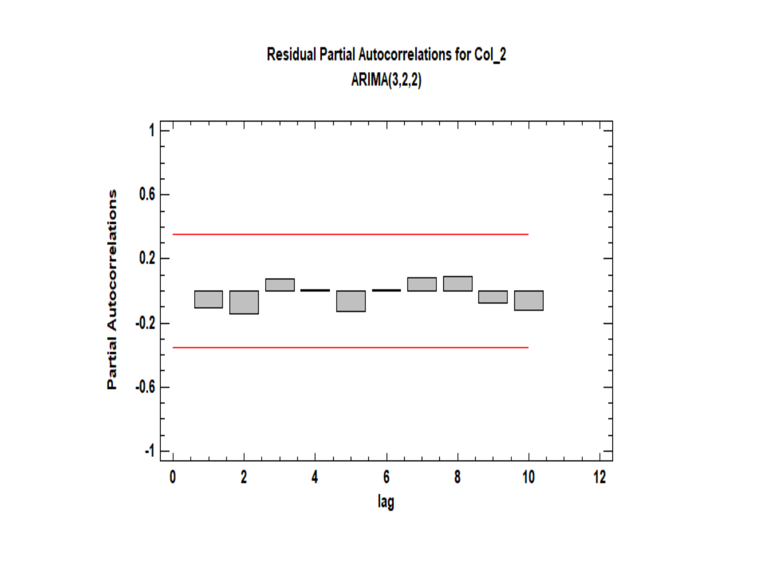

FIGURE 2. Residuals PACF Plot for ARIMA $(3,2,2)$

TABLE 5. One step ahead forecasts and residuals for the LEB data (1980-2012)

Period	Data	forecasted	residual
2013	65.927	65.9268	0.0002
2014	66.139	66.142	-0.0030
2015	66.322	66.3424	-0.0204
2016	66.481	66.5468	-0.0658
2017	66.630	66.776	-0.141

4. CONCLUSION

In this study, the LEB in Pakistan for the time 2017-2025 was forecasted with high accuracy by using the ARIMA model among all other counterparts' models. The best model is selected based on model selection criteria i.e. AIC and SBIC. On the model selection criteria, we found the ARIMA $(3,2,2)$ is the finest model for forecasting the LEB of Pakistan. To checks the assumptions of the best-fitted model we utilized different run tests to identify the randomness of the residuals, the JB test for justifying the normality of residuals, and the ACF plot autocorrelation. The best-fitted model full fill all require

TABLE 6. Life Expectancy at Birth forecasts (in percentage)

Years	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
2018	67.05	66.87	67.23
2019	67.36	67.11	67.61
2020	67.70	67.39	68.02
2021	68.05	67.66	68.44
2022	68.37	67.91	68.83
2023	68.64	68.11	69.16
2024	68.83	68.24	69.42
2025	68.97	68.31	69.63

assumptions. On the basis of ARIMA $(3,2,2)$ the LEB would be 67.05 and 68.97 in 2020 and 2025, respectively.

Time series analysis is a useful tool to find the future behavior of the country. It's effective for policymakers and planning for the upcoming years. Awareness of people, good health facilities, and a clean environment would increase the LEB in Pakistan. We hope that these time series results provide the baseline for government to make new policies to increase the LEB.

REFERENCES

[1] World Bank (2020). Life expectancy at birth, total (years). Pakistan https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=PK.

[2] Guillen, M., Vidiella-i-Anguera, A. (2005). Forecasting Spanish natural life expectancy. *Risk Analysis: An International Journal*, **25**(5), 1161-1170.

[3] Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., & Ludwig, D. S. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine*, **352**(11), 1138-1145.

[4] Skiadas, C. H. (2011). Life expectancy at birth, estimates and forecasts in the Netherlands (females). *arXiv preprint arXiv:1112.0796*.

[5] Torri, T., Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, **28**(2), 519-531.

[6] Le Bourg, E. (2012). Forecasting continuously increasing life expectancy: what implications?. *Ageing research reviews*, **11**(2), 325-328.

[7] Bennett, J. E., Li, G., Foreman, K., Best, N., Kontis, V., Pearson, C., Ezzati, M. (2015). The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *The Lancet*, **386**(9989), 163-170.

[8] Shang, H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *Annals of Applied Statistics*, **10**(3), 1639-1672.

[9] Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K., Ezzati, M. (2017). Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *The Lancet*, **389**(10076), 1323-1335.

[10] Pascariu, M. D., Canudas-Romo, V., Vaupel, J. W. (2018). The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*, **78**, 339-350.

[11] Wu, R., Wang, B. (2018). Gaussian process regression method for forecasting of mortality rates. *Neurocomputing*, **316**, 232-239.

[12] Shair, S. N., Zolkifi, N. A., Zulkefi, N. F., Murad, A. (2019). A Functional Data Approach to the Estimation of Mortality and Life Expectancy at Birth in Developing Countries. *Pertanika Journal of Science & Technology*, **27**(2), 797-814.

[13] Nigri, A., Levantesi, S., Marino, M. (2021). Life expectancy and lifespan disparity forecasting: a long short-term memory approach. *Scandinavian Actuarial Journal*, **2**, 110-133.

[14] Perla, F., Richman, R., Scognamiglio, S., Wthrich, M. V. (2021). Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*,1-27.

[15] Amir, M. W., Bibi, A., Akhtar, N., Raza, Z. (2021). Modeling and Forecasting of Gross Domestic Product Percentage Share of Education Sector: A Statistical Study in Pakistan. *Transactions in Mathematical and Computational Sciences*, **1**(1), 75-84.

[16] Wei, W. W. S. (2006). Time series analysis. *Addison Wesley Publishing Company, New York*.

[17] Cooray, T. M. J.A. (2008). Applied time series analysis and forecasting. *Narosa Publishing House Pvt. Ltd*.

[18] elik, S., Karadas, K., Eyduran, E. (2017). Forecasting the production of Groundnut in Turkey using Arima model. *The Journal of Animal & Plant Sciences*, **27**(3), 920-928.

[19] Amir, M. W., Amin, M., Nazir, H. Z. (2021). Statistical Modeling and Forecasting for Onion Production of Pakistan. *Journal of Agricultural Research*, **59**(2), 171-176.

[20] Chatfield, C (1995). The Analysis of Time Series an Introduction. *5th Ed. Chapman & Halljcrc*.

[21] Gujarati, D. N (2009). Basic econometrics. *Tata McGraw-Hill Education*.