Exploring Monthly Births Trend in Sri Lanka

Abstract

Monthly births are important to a country as they directly influence its population size and demographic structure, economy, and healthcare system. As Sri Lanka is a developing country, analyzing monthly births can be introduced as a major aspect of resource management. The number of monthly births during 2012 – 2021 was obtained from the Register General's Department of Sri Lanka. There has been no attempt to access the forecasting monthly birth count so far. The main objective of the study was to predict a suitable model and forecast future monthly births. In this study, we predicted seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters methods and did a comparison based on the residuals to find the best model using R programming. According to the results, the best prediction model was selected based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). By the comparison of SARIMA and Holt-Winters methods, the SARIMA (4,1,1)(2,0,0)[12] model was the best-fitted statistical model for this monthly births data. The predicted results for 2022 from this model were compared with the actual monthly birth count to decide the accuracy of the prediction.

Introduction

The study of monthly births is an important factor for making policies, healthcare planning, and other main resource management in a country. Examining monthly births in Sri Lanka, a country known for its rich cultural heritage and multicultural population, offers important insights into the nation's reproductive health environment, healthcare system, and socioeconomic influences on family planning.

Jorge Miguel Bravo and Edviges Coelho evaluate the forecasting performance of alternative linear and non-linear time series methods to birth and death monthly forecasting at the sub-national level by using seasonal ARIMA, HoltWinters and State Space models. For both male and female subpopulation births and deaths, simple and weighted average predicting performance for the three models was roughly identical; however, State Space models performed somewhat better for births and seasonal ARIMA for deaths [1].

By using seasonal ARIMA time series analysis techniques, Kenneth C. Land and David Cantor produce final models that contain significant second-order autoregressive components as well as seasonal moving-average components for both series for the U.S. monthly birth and death rates for January 1950–December 1978 [2].

Buah Ahoba Masha, Zigli David Delali, and Annan Wobir Reuben have found a model and forecast the monthly Birth rate in Ghana using the Box-Jenkins method.ARIMA (1,1,1) model was the most suitable model with the least normalized Bayesian Information Criterion (BIC), and Akaike Information Criterion(AIC) values amidst the model test in this research. In their conclusions, there was a negative relationship between the Birth rate and months [3].

Materials & Methods

The Monthly births data were collected from the Registrar General's Department in Sri Lanka in the time period from 2012 to 2021.The data was analyzed using R software version 4.2.1 .

Holt Winter's Exponential Smoothing

Holt Winter's methods are used, when the data series has a trend and seasonal pattern. This method is based on three smoothing equations - one for the level, one for the trend and one for the seasonality. According to the pattern of the seasonality, there are two formulations in Holt Winter's Exponential Smoothing such as additive and multiplicative.

The additive method is specified as:

$$L_t = \alpha\ (y_t - S_{t-p})\ +\ (1-\alpha)\ (L_{t-1} - T_{t-1})\ \ 0 \leq\ \alpha \leq 1\ (Level\ equation) \tag{1}$$

$$T_t =\ \beta\ (L_t -\ L_{t-1})\ +\ (1-\ \beta)\ T_{t-1} \qquad (Trend\ \ estimating\ equation) \tag{2}$$

$$S_t =\ \gamma\ (y_t -\ L_t)\ +\ (1-\ \gamma)\ S_{t-p} \qquad (Seasonality\ updating\ equation\ ) \tag{3}$$

$$F_{t-m} =\ L_t\ + mT_t\ +\ S_{t+m-p} \qquad (Forecast\ equation\ ) \tag{4}$$

Where smoothing parameters α , β and γ; F_{t+m} is the forecast for m periods ahead at time t.

The Holt-Winters' multiplicative method is defined as:

$$L_t = \alpha\ \frac{y_t}{S_{t-p}}\ +\ (1-\alpha)\ (L_{t-1} - T_{t-1}) \qquad (Level\ equation) \tag{5}$$

$$T_t =\ \beta\ (L_t -\ L_{t-1})\ +\ (1-\ \beta)\ T_{t-1} \qquad (Trend\ \ estimating\ equation) \tag{6}$$

$$S_t =\ \gamma\ \frac{y_t}{L_t}\ +\ (1-\ \gamma)\ S_{t-p} \qquad (Seasonality\ updating\ equation\ ) \tag{7}$$

$$F_{t-m} = \ (L_t \ + mT_t \ ) \ S_{t+m-p} \qquad (Forecast\ equation\ ) \tag{8}$$

Where smoothing parameters α , β and γ; F_{t+m} is the forecast for m periods ahead at time t.

Seasonal ARIMA Model

The seasonal Autoregressive Integrated Moving Average (ARIMA) model supports to model both trend and seasonal components of a time series and used to forecasting. The model includes new parameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

$$\phi_P\left(B^S\right)\phi\left(B\right)\nabla_S^D\ \nabla^d x_t \ \ = \ \delta + \Theta_Q(B^s)\theta(B)w_t \tag{9}$$

where w_t denotes the Gaussian white noise process. The general model can be expressed as ARIMA(p,d,q) $\times$ $(P, D, Q)_s$, where the ordinary autoregressive (AR) and moving average (MA) components are represented by polynomials$\phi\left(B\right)$ and $\theta\left(B\right)$ of orders p and q, respectively, the seasonal AR and MA components are denoted by $\phi_P\left(B^S\right)$ and $\Theta_Q\left(B^S\right)$ of orders P and Q, respectively. The non-seasonal and seasonal difference components are represented by $\nabla^d \ = \ (1-B)^d$ and $\nabla^D \ = \ \left(B^S\right)^D$, respectively. The seasonal period defines the number of observations that make up a seasonal cycle (e.g., s= 12 for monthly observations).

Results and Discussion

According to time series plot we identified that there was a decreasing trend and additive seasonal pattern. To verify the patterns and trends, the dataset was decomposed. So, we applied Holt Winter's exponential smoothing to dataset. We analyze the stationary of the series using ADF (Augmented Dickey Fuller) test to check whether or not a seasonal difference is needed to produce a roughly stationary series. According to the results of ADF test (p-value $<$ 0.01), the series is stationary. Then for both Holt Winter's exponential smoothing and ARIMA models and analyzed residuals to evaluate forecast accuracy using different criteria (the Root-mean-square error (RMSE), the Mean Absolute Error (MAE), the Monthly Percentage Error (MPE), the Mean Absolute Percent Error (MAPE), the Mean Absolute Scaled Error (MASE)).

The results are shown in following table:

Error	Holt Winter's	SARIMA
RMSE	2684.146	2257.275
MAE	-2.534114	-2.02103
MAPE	7.109025	6.475313
MASE	0.8304944	0.7829551

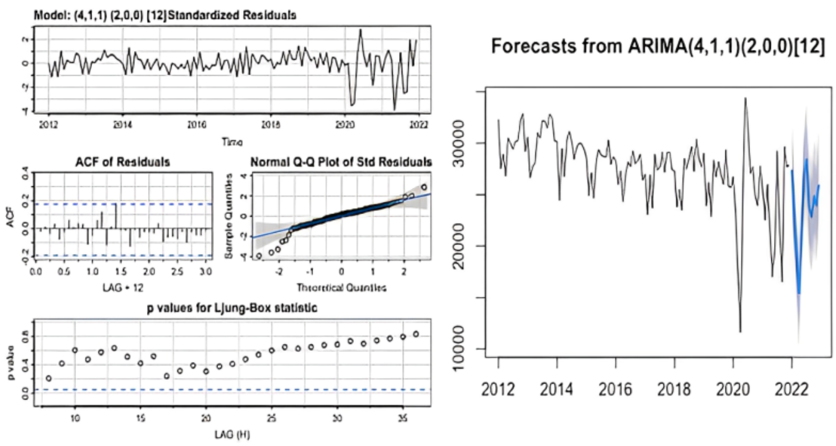

Figure 1: Residual analysis of SARIMA(4,1,1)(2,0,0)[12] model and the Forecated values

Conclusion

By comparing the residuals of Holt Winter's exponential smoothing and SARIMA models, we select SARIMA (4,1,1)(2,0,0)[12] model as the best model for forecasting monthly births in Sri Lanka. The model equation is:

$$\begin{aligned} Y_t = 1.5094Y_{t-1} - 0.9535Y_{t-2} + 0.7109Y_{t-3} - 0.4609Y_{t-4} + 0.1941Y_{t-5} + 0.2717Y_{t-12} \\ - 0.27117Y_{t-24} + 0.9385e_{t-1} + e_t \end{aligned}$$

Forecasted values for twelve months in year 2022 ; 27340.26, 22240.94, 18739.57, 15442.86, 21758.35, 26903.1, 28421.89, 23870.32, 22826.66, 24864.93, 23884.93, 23884.77 and 25875.88 respectively.

References

[1] J. M. Bravo and E. Coelho. Forecasting subnational monthly births and deaths using seasonal time series methods. *Evidence-based territorial policymaking: formulation, implementation and evaluation of policy*, pages 1079–1088, 2019.

[2] K. C. Land and D. Cantor. Arima models of seasonal variation in us birth and death rates. *Demography*, 20:541–568, 1983.

[3] B. A. Masha, Z. D. Delali, and A. W. Reuben. Time series analysis of monthly birth rate in ghana. *International Journal of Applied Science and Mathematics*, 7(5):2394–2894, 2020.