Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 32, 2025, page 1-10

Time Series Forecasting on Live Births in Malaysia

Ng Jie Han[1], Nur Syazwin Mansor[1]*, Adina Najwa Kamarudin[2]
[1]Department of Mathematical Sciences, Faculty of Science,
Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia.
[2]Institute of Mathematical Sciences, Faculty of Science,
Universiti Malaya, Kuala Lumpur
*Corresponding author: nursyazwin@utm.my

Abstract
Live birth trends are a key demographic indicator that reflect population dynamics and play a crucial role in long-term planning for sectors such as healthcare, education, and economic development. Accurate forecasting of live births enables policymakers to allocate resources effectively and anticipate future societal needs. This models and forecasts the number of daily live births in Malaysia using data from the year 2022. Two time series forecasting methods were applied: the Seasonal Autoregressive Integrated Moving Average (SARIMA) Model and Holt-Winters' Exponential Smoothing Model. Model performance was evaluated using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The Holt-Winters method outperformed the SARIMA model, yielding the lowest RMSE and MAPE values and was thus selected for forecasting the daily live births for the year 2023. The forecast revealed a consistent seasonal pattern with a slight upward trend, indicating stable weekly birth cycles throughout the year. These findings provide valuable insights into Malaysia's live birth trends and demonstrates the effectiveness of time series forecasting models for demographic research and national policy planning.
Keywords: Live Births; Forecasting; SARIMA; Holt-Winters'

1. Introduction

Live birth trends play a vital role in shaping the demographic profile and economic future of a nation. In Malaysia, the declining fertility rate has become a growing concern, as it may lead to long-term challenges such as an aging population, labor shortages and reduced economic productivity. According to the Department of Statistics Malaysia (DOSM), the country recorded a fertility rate of only 1.6 children per woman in 2022, the lowest in the past five decades. Understanding and forecasting live birth patterns is essential for policymakers, health planners and educators to prepare for and allocate resources efficiently.

Various socio-economic factors such as urbanization, lifestyle changes, and delayed marriages have contributed to the reduction in Malaysia's live birth numbers. While qualitative insights into these issues are well-documented, there is a lack of effective quantitative forecasting models that can capture daily live birth trends in Malaysia. The ability to accurately forecast future live births is crucial for national planning, particularly in sectors such as healthcare infrastructure, maternity services, childcare, and education.

This study aims to develop a time series forecasting model to predict the number of live births in Malaysia using daily birth data from the year 2022. Two forecasting methods are utilized: Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters' Exponential Smoothing. The objectives are to analyze the trend and seasonality in the live birth data, determine the most suitable forecasting model, and estimate the number of live births for 2023. By evaluating the models using accuracy metrics such as RMSE and MAPE, this research offers data-driven insights into Malaysia's demographic shifts and supports informed decision-making for sustainable development.

2. Literature Review

The decline in Malaysia's live birth reflects broader global trends influenced by a combination of socio-economic, biological and environmental factors. Understanding these drivers, along with suitable forecasting methods is crucial for projecting future population patterns and guiding effective policy development.

Malaysia has seen a sharp decrease in live births over recent decades. According to the Department of Statistics Malaysia (DOSM), the total fertility rate dropped to 1.6 children per woman in 2022 which is the lowest in 50 years. This decline mirrors global patterns observed in both developed and developing countries, such as the U.S. and China, where shifts in economic priorities, urbanization, and evolving family structures contribute to lower fertility. For example, Zhang [1] found that policies like China's one-child policy and changing societal norms have led to long-term demographic impacts, including reduced family size and declining birth rates. In Malaysia, increasing female labor force participation, delayed marriages, and rising living costs further contribute to this trend [2]. The average age of marriage in Malaysia increased from 24.7 years in 1990 to 28.9 in 2022, which correlates with lower fertility levels [3].

Despite the importance of understanding live births patterns, many existing studies focus primarily on explaining factors rather than forecasting future trends. This gap highlights the need for effective time series forecasting techniques to support data-driven decision-making. Several studies have successfully applied models such as ARIMA and Holt's Exponential Smoothing to forecast fertility rates in various countries. For example, Caleiro [4] used time series methods to predict birth rates in Portugal, while Yang et al. [5] applied ARIMA models in Australia, demonstrating the model's capability to account for seasonality and trends.

Holt's Exponential Smoothing and its extended forms, such as the Holt-Winters method, have also shown reliable performance in forecasting demographic data with consistent patterns. Kumar and Meena [6] compared Holt's and ARIMA models in forecasting India's fertility rate, revealing ARIMA's slight superiority in accuracy, while still acknowledging Holt's suitability for smoother trends. Similarly, Wang [7] evaluated both ARIMA and Exponential Smoothing State Space Models (ETS) for China's birth rates, ultimately favoring ARIMA due to its strength in handling complex seasonality.

In summary, various time series forecasting models have been proven effective for demographic prediction. While ARIMA remains a robust and versatile choice, Holt-Winters' method offers practical advantages for data with clear seasonal components. This study applies both SARIMA and Holt-Winters' models to Malaysian daily birth data, aiming to determine the most accurate forecasting approach for projecting future birth trends.

3. Methodology

3.1. Data Description

The aim of this study is to forecast the number of live births in Malaysia for 2023, using daily live births data from 2022 obtained from the Department of Statistics Malaysia (DOSM) through the official Malaysian government data portal, yielding a total of 365 data points for the year.

3.2. SARIMA Model

The SARIMA approach is an extension of ARIMA that handles seasonal components in time series data. Unlike ARIMA, SARIMA adds parameters to model seasonal patterns, making it suitable for datasets like live births, which often exhibit yearly cycles or periodic trends. SARIMA is expressed as ARIMA$(p,d,q)(P,D,Q)[m]$, where the seasonal components (P,D,Q) account for repeating patterns within the data. P represents seasonal autoregressive order, D represents seasonal differencing order while Q is seasonal moving average order and m represents seasonal period. If the data exhibits seasonal patterns, use seasonal differencing formula $Y_t' = Y_t - Y_{t-365}$ to remove seasonality, leaving only the trend and residual components for modelling. Then, plot the ACF and PACF plot to determine spikes at multiples of the seasonal period (e.g. 365 for yearly) and detect seasonal lags by observing significant correlations respectively. Apart from that, use statistical software like R programming to fit the SARIMA model. Also, check residuals for patterns to validate the model and evaluate AIC and BIC to compare model performance (same with steps in ARIMA model). Finally, Generate predictions while accounting for seasonal cycles. SARIMA forecasts not only trends but also repeating seasonal patterns based on historical cycles. Visualize predictions along with confidence intervals to understand variations.

3.3. Holt-Winters' Exponential Smoothing

Holt-Winters' Exponential Smoothing, also known as Triple Exponential Smoothing, is an advanced time series forecasting method designed to handle data with trend and seasonality components. Unlike Simple Exponential Smoothing (SES), which is suitable for data without trends or seasonality, Triple Exponential Smoothing incorporates additional parameters to account for seasonal variations and long-term trends, making it particularly effective for datasets like Malaysia's daily live births in 2022, which exhibits clear seasonal patterns.

This method operates through two main models: additive model and multiplicative model. The additive model performs well for data with constant seasonal variations while the multiplicative model is suitable for proportional seasonal variations which means the fluctuations increase or decrease relative to the trend. In this study, Triple Exponential Smoothing with additive model will be applied.

Formally, Holt-Winters' smoothing approach employs three smoothing constants, denoted by α, β and γ. The Holt-Winters' smoothing model is as follows:

$$F_{t+m} = L_t + b_t m + S_{t+m-s} \tag{1}$$

Where:

Level series:

$$L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{2}$$

Trend estimate:

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{3}$$

Seasonality factor

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s} \tag{4}$$

Where α = level smoothing parameter with the interval of 0 to 1, β = trend smoothing parameter with the interval of 0 to 1, γ = seasonal smoothing parameter with the interval of 0 to 1 and m = seasonal period.

The procedures for the Holt-Winters' Exponential Smoothing Method are as follows: Firstly, obtain initial values for the level L_0, the growth rate b_0 and the seasonals factors $S_1, S_2, S_3, \dots, S_s$ by setting $L_s = \frac{1}{s}(y_1 + y_2 + \cdots + y_s), b_s = \frac{1}{s}\left(\frac{y_{s+1}-y_1}{s} + \frac{y_{s+2}-y_2}{s} + \cdots + \frac{y_{s+s}-y_s}{s}\right)$ and $S_1 = y_1 - L_s, S_2 = y_2 - L_s, \dots, S_s = y_s - L_s$. After that, calculate a point forecast of F_{t+1} using the initial values $F_{t+m} = L_t + b_t m + S_{t+m-s}$. Then, update the estimates L_t, b_t and S_t by using the formula (3.7), (3.8) and (3.9). Lastly, Find the most suitable combination of α, β and γ that minimizes the error measurements.

3.4. Forecasting Performance Evaluation

Statistics that are used to inspect the accuracy of the forecasting model include the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAPE is a measurement to evaluate the forecasting accuracy of the prediction model in percentage, and it is used as a loss function for regression analysis. The MAPE can be calculated from the following formula:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100 \tag{5}$$

Where $\hat{y}_t$ is the predicted value, y_t is the actual value at time t and N is the size of the sample.

On the other hand, the root mean square error (RMSE) is known as the standard deviation of the prediction errors (residual). RMSE indicates that the absolute fit to the forecast model with data. The RMSE can be calculated from the following formula:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y_t})^2} \quad (6)$$

Where $\hat{y_t}$ is the predicted value, y_t is the actual value at time t and N is the size of the sample. An outperformed model should comprise the lowest MAPE and RMSE.

4. Results and discussion

4.1. Trend Analysis

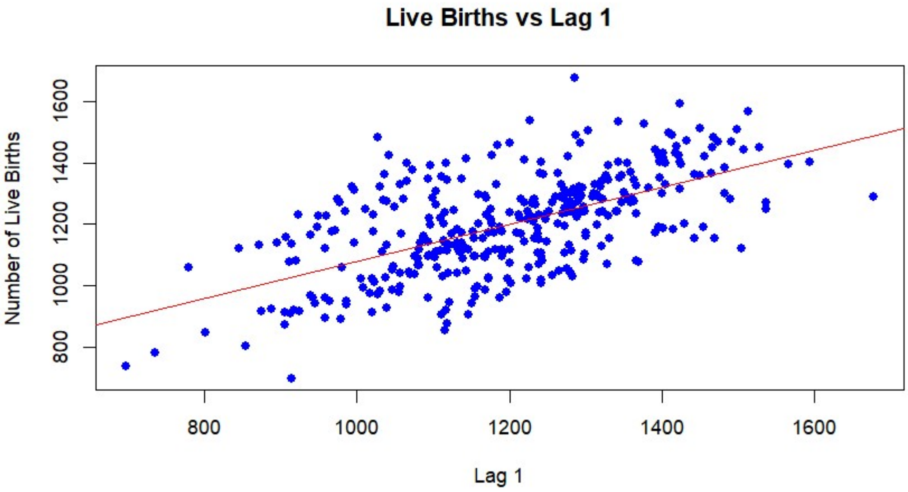

Figure 1 Lag Plot of Live Births in 2022

The lag plot in Figure 4.1 shows a dense clustering of points forming an elongated shape, suggesting a moderate to strong linear relationship. This suggests that the number of live births on a given day tends to be somewhat correlated with the number born the previous day, indicating potential autocorrelation in the time series. However, it may not be a perfect straight line, implying that other factors or randomness also influence daily live births.

4.2. Time Series Analysis

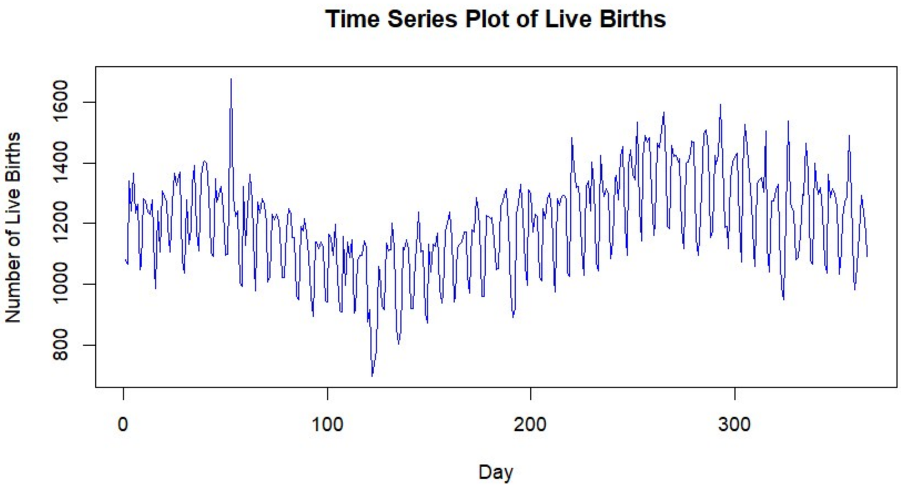

Figure 2 Time Series Plot of Live Births in Malaysia in year 2022

The time series plot of Malaysia's daily live births in 2022 is visualized as in Figure 2. Based on the time series plot, there are noticeable fluctuations over the year. The number of live births range between around 800 and 1600 per day and exist several peaks and troughs across the timeline. A significant spike can be identified around the 60th observation, possibly indicating specific events or holidays influencing live births. Following this, there is a gradual decline until around the 120th observation, after which the data appears to stabilize before showing an increasing trend. The trend reaches a higher level between observations 200 and 300, indicating a seasonal pattern with noticeable periodic variations.

4.3. Time Series Forecasting

In this study, the dataset is divided into two parts to ensure effective model evaluation. The first 294 observations represent the training set (in-sample), which is used to develop and fit the forecasting model. The remaining 71 observations form the test set (out-sample), which serves to evaluate the model's forecasting performance on unseen data. This approach helps to validate the model's accuracy and its ability to generalize to future values.

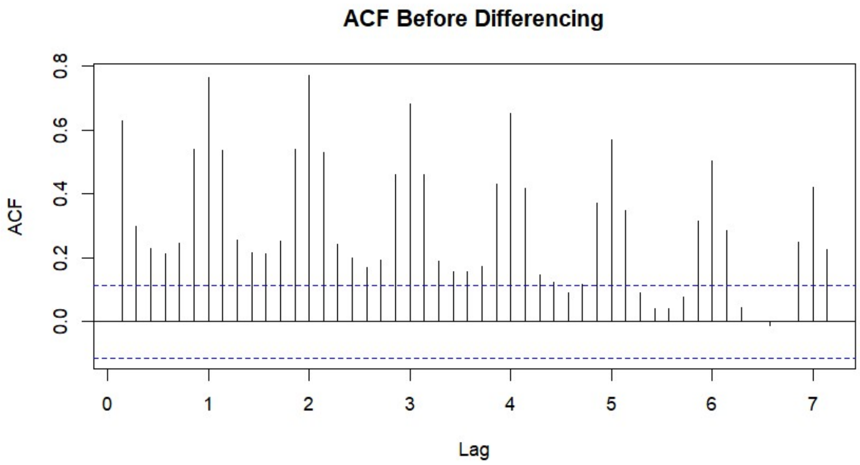

Figure 3 ACF plot before differencing

Based on Figure 3, the ACF plot shows autocorrelations decaying slowly over many lags instead of a sharp cutoff. This persistent autocorrelation at higher lags (extending beyond 5-10 lags) suggests that the series exhibits either trend or seasonality, causing the mean or variance to change over time. This indicates that the series is not stationary.

```
        Augmented Dickey-Fuller Test

data:  train_data
Dickey-Fuller = -1.6115, Lag order = 6, p-value = 0.7399
alternative hypothesis: stationary
```

Figure 4 ADF test before differencing

The ADF test is used to justify the stationarity for the time series data as shown in Figure 4. Null hypothesis state that the time series is stationary while alternative hypothesis means the time series is non-stationary. The result shows a p-value of 0.7399, which is greater than 0.05 significance level. Thus, fail to reject null hypothesis at 5% significance level, indicating that the time series is non-stationary. The time series should be differenced to employ the SARIMA model.

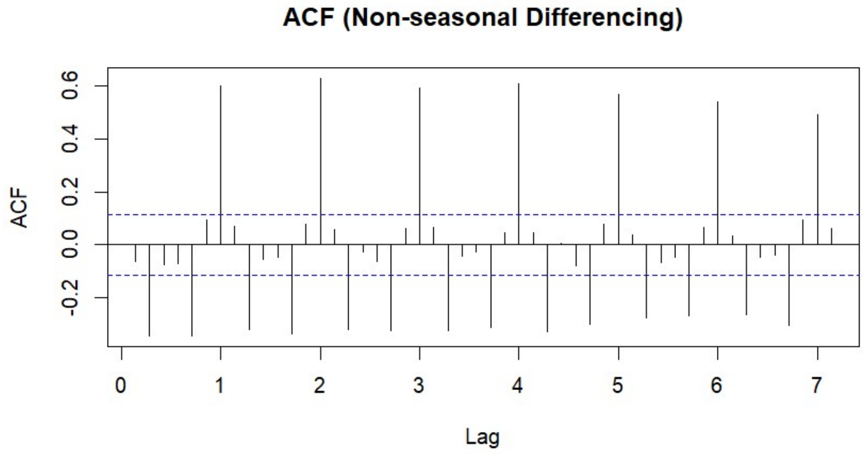

Figure 5 ACF of differenced series in order, $d = 1$

Based on the ACF plot in Figure 5, most of the autocorrelation values fall within the 95% confidence limits, represented by the red lines, suggesting that the series has been mostly transformed into a stationary series. However, a few spikes exceed the confidence bounds, indicating some remaining autocorrelation at specific lags. This is typical for time series data after differencing, especially if there are seasonal or periodic components that have not been fully removed.

```
        Augmented Dickey-Fuller Test

data:  diff_train
Dickey-Fuller = -11.751, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Figure 6 ADF test of first differencing

The result in Figure 6 shows a p-value of 0.01, which is smaller than 0.05 significance level. Thus, reject the null hypothesis at 5% significance level, indicating that the trend component has been removed.

Other than that, seasonality is one of the important considerations in times series analysis. Before modelling the data to the forecasting models, the seasonality of the dataset should be examined.

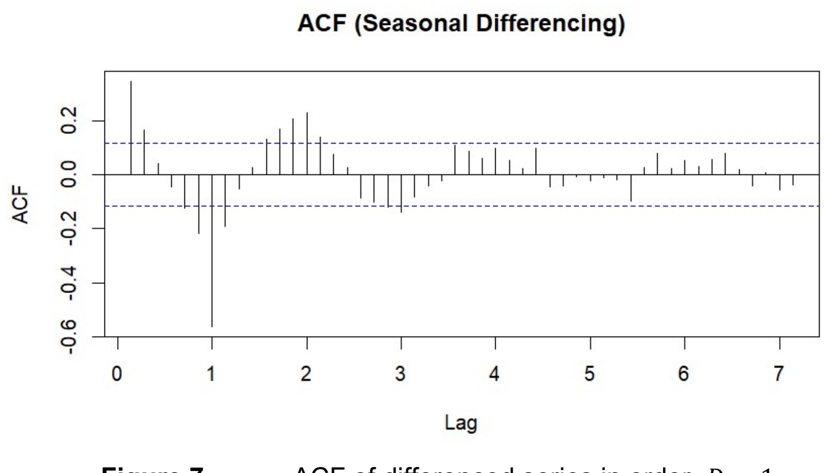

Figure 7 ACF of differenced series in order, $D = 1$

As shown in Figure 7, most of the autocorrelation values now fall within the 5% significance limits. This suggests that the data is approaching stationarity, as there are no significant autocorrelations beyond lag 0. Additionally, the absence of oscillatory patterns or periodic peaks in the plot indicates that the seasonal effects have been successfully removed or minimized.

```
        Augmented Dickey-Fuller Test

data:  seasonal_diff_train
Dickey-Fuller = -13.342, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Figure 8 ADF test of second differencing

Figure 8 shows that the stationarity of the second differencing data is examined using the ADF test. The result shows a p-value of 0 which is smaller than 0.05, thus null hypothesis is rejected indicating that the first differencing data appears to be stationary.

4.3. Forecasting using SARIMA

SARIMA$(1,1,1)(1,1,1)_7$ is chosen since it has the lowest AICc value and diagnostic checking is applied to ensure the model is adequate to forecast the result with minimum prediction error. The residuals must satisfy the properties of zero mean, independence, constant variances, and normality as in Figure 9, 10, 11 and 12.

```
> print(t_test_results)
     Coefficient  Std_Error    t_value      p_value
ar1    0.2890911 0.06687198   4.323053 2.120122e-05
ma1   -0.8768416 0.03125891 -28.050930 0.000000e+00
sar1  -0.1144820 0.07377604  -1.551750 1.218164e-01
sma1  -0.8884168 0.06260372 -14.191119 0.000000e+00
```

Figure 9 t-test of SARIMA$(1,1,1)(1,1,1)_7$

```
p-value: 0.8336
```

Figure 10 Ljung-Box Test of SARIMA$(1,1,1)(1,1,1)_7$

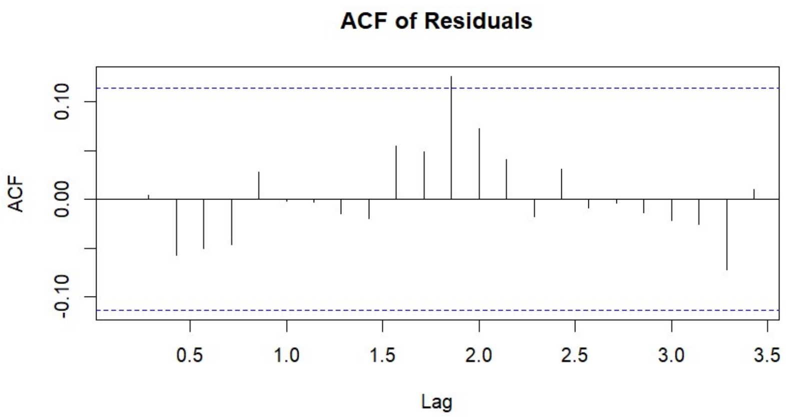

Figure 11 ACF Plot of SARIMA$(1,1,1)(1,1,1)_7$ Residuals

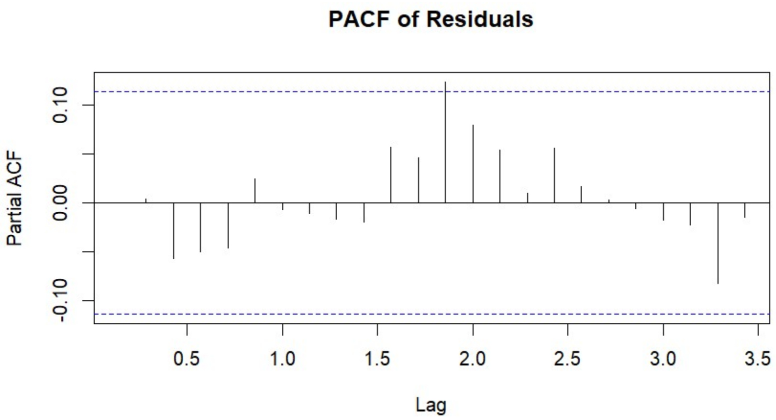

Figure 12 PACF Plot of SARIMA$(1,1,1)(1,1,1)_7$ Residuals

Overall, the residuals of SARIMA$(1,1,1)(1,1,1)_7$ are white noise since it satisfies all the major assumptions of residuals. The finalised equation of this forecasting model after rearrangement can be formulated as:

$$y_t' = 0.2891y_{t-1}' - 0.1145y_{t-7}' + 0.0331y_{t-8}' + \varepsilon_t - 0.8768\varepsilon_{t-1} - 0.8884\varepsilon_{t-7} + 0.7791\varepsilon_{t-8} \quad (7)$$

Where:

$$y_t' = (1 - B)(1 - B^7)y_t \quad (8)$$

Figure 13 illustrates the forecasting trend by using SARIMA model.

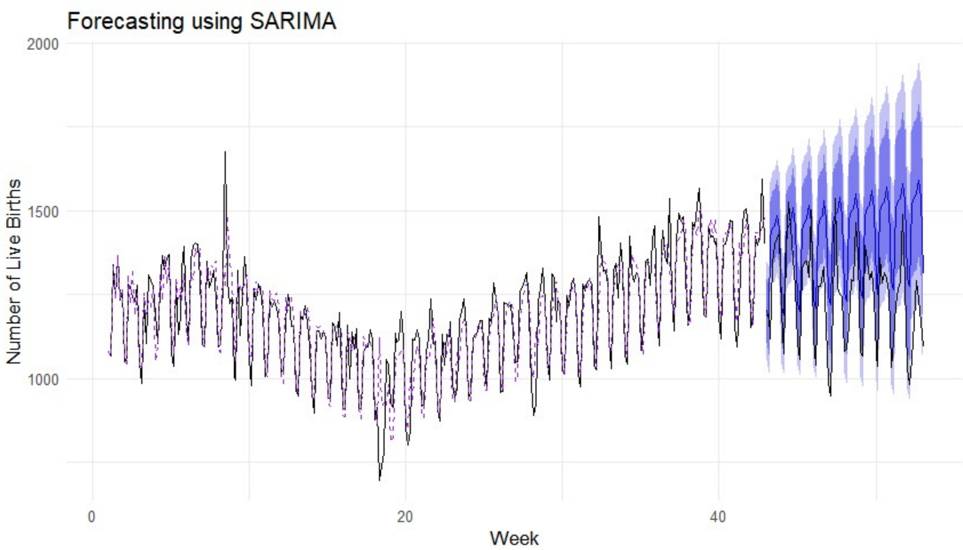

Figure 13 Forecasting Plot using SARIMA

4.4. Forecasting using Holt-Winters Exponential Smoothing

The optimal parameters chosen will imply the fitted value of the model closer to the actual data in training samples. Based on Figure 4.17, the equation of the Holt-Winters' model can be expressed by using the general formula in Equation 9:

$$F_{t+m} = L_t + b_t m + S_{t-m+s} \quad (9)$$

With

$$L_t = 0.2472(y_t - S_{t-s}) + (1 - 0.2472)(L_{t-1} + b_{t-1}) \quad (10)$$

$$b_t = 0.0001(L_t - L_{t-1}) + (1 - 0.0001)b_{t-1} \quad (11)$$

$$S_t = 0.0004(y_t - L_t) + (1 - 0.0004)S_{t-s} \quad (12)$$

Figure 14 illustrates the forecasting trend by using Holt-Winters' model.

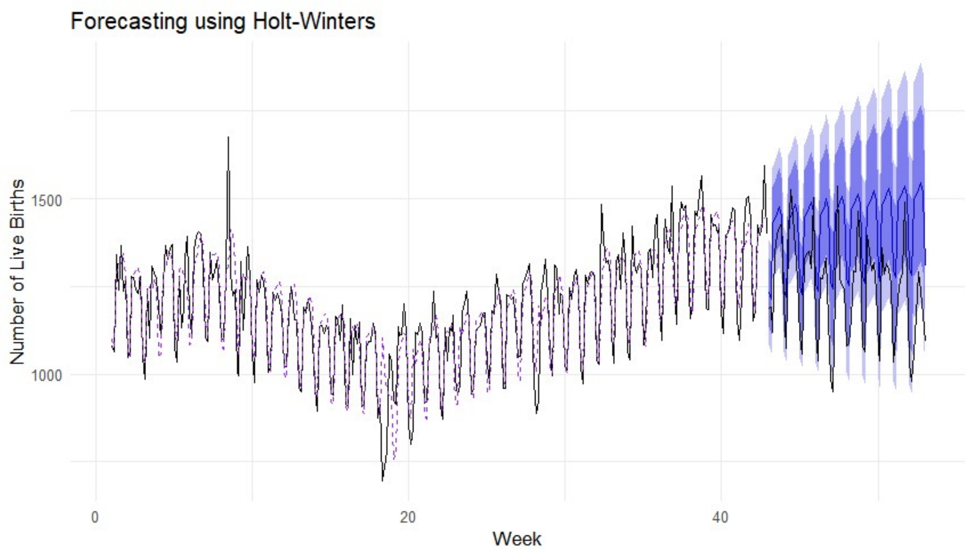

Figure 14 Forecasting Plot using Holt-Winters

4.5. Forecasting Performance Evaluation

Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters' Method have been used to model the live births in 2022. The forecast of number of live births in 2023 are completed by using these two models.

Table 1: Forecast Accuracy of Models

Model	Modelling		Forecasting	
	MAPE (%)	RMSE	MAPE (%)	RMSE
SARIMA	**4.3613**	**73.2179**	15.7255	210.8760
Holt-Winters'	4.5878	74.6039	**15.3881**	**203.4570**

According to Table 1, the best modelling model is SARIMA while the best forecasting model is Holt-Winters'. In modelling, SARIMA achieving slightly better accuracy compared to Holt-Winters'. This suggests that SARIMA had a marginally better fit to the training data. However, when evaluated on the test data, Holt-Winters' Exponential Smoothing model outperformed SARIMA, producing a lower MAPE compared to SARIMA's. These results indicate that Holt-Winters' model generalizes better and provides more reliable forecasts for future live birth trends.

Conclusion

This study applied time series forecasting methods, specifically SARIMA and Holt-Winters' Exponential Smoothing, to predict the number of live births in Malaysia using daily data from 2022. After thorough data exploration, decomposition, and model evaluation, the Holt-Winters model emerged as the most accurate, with the lowest RMSE and MAPE values. Its ability to capture both trend and seasonality made it the preferred model for forecasting births in 2023. The forecast revealed a consistent weekly seasonal pattern with a slight upward trend, possibly reflecting post-pandemic recovery and sociocultural influences. These findings underscore the importance of accurate birth forecasting in supporting effective planning for healthcare, education and public policy.

Based on the forecast plot using Holt-Winters model in Figure 14, the forecasted number of live births in Malaysia for early 2023 demonstrates a continuation of the weekly seasonal pattern observed in 2022, along with a gradual upward trend. This consistent seasonality, indicated by regular fluctuations, reflects typical weekly cycles in birth occurrences, which could be influenced by hospital scheduling patterns, cultural preferences, or administrative reporting routines. The slight increasing trend in the forecast may suggest a potential rise in number of live births due to various demographic and social factors, such as improved post-pandemic stability, delayed family planning resuming after COVID-19 or growing populations in certain regions.

Acknowledgement
I wish to express my sincere gratitude to all who have contributed throughout the course of this work.

References

[1] Zhang, J. (2017). *The Evolution of China's One-Child Policy and Its Impact on Fertility Rates*. Journal of Population Economics, 30(2), 651–672.

[2] Siah, C.L., & Lee, M. (2015). *Female Labour Force Participation and Its Impact on Fertility Rate in Malaysia*. Malaysian Journal of Economic Studies, 52(1), 31–47.

[3] Bernama. (2024). *Fertility Rate Decline in Malaysia Due to Late Marriages*. Department of Statistics Malaysia. Retrieved from https://www.dosm.gov.my

[4] Caleiro, A. (2010). *Forecasting the Number of Births in Portugal: A Time Series Approach*. Journal of Demographic Research, 12(1), 14–24.

[5] Yang, S., Shang, H., & Raymer, J. (2022). *Forecasting Australian Fertility Using ARIMA Models*. Population Studies, 76(1), 102–118.

[6] Kumar, A., & Meena, R. (2024). *Comparative Analysis of Holt's and ARIMA Models for Forecasting India's Total Fertility Rate*. International Journal of Statistics and Demography, 8(2), 55–66.

[7] Wang, Y. (2023). *A Comparative Study on ARIMA and ETS Models in Predicting China's Birth Rate*. Asian Population Journal, 19(1), 87–103.