Proceedings of Science and Mathematics
Faculty of science, universiti teknologi malaysia
https://science.utm.my/procscimath
Volume 21 (2024) 142-156

Time Series Forecasting of Crude Birth Rate in Malaysia

Lynlie E Lian, Muhammad Fauzee Hamdan*
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: mfauzee@utm.my

Abstract
Malaysia's birth rate has drastically decreased over the past century, particularly in the last several years. The decline in fertility was primarily caused by postponed childbirth and low fertility intentions. The issue of aging in our nation is becoming more and more pressing as time goes on. As a result, the increasing of birth rate has taken precedence in both society and the government. In this study, ARIMA and ETS model was used in Minitab to model and forecast Malaysia's fertility rate. Additionally, the data from the Malaysian National Statistics Office was used and the historical fertility rate in Malaysia from 2001 to 2002 was chosen. The goal of this study is to determine which models is the most effective model for short-term projections of Malaysia's future birth rate. Hence, to create a stationary birth rate time series, the data was first transformed. Then, the ARIMA and ETS models were used, respectively, to determine which model was more accurate in predicting Malaysia's short-term birth rate over the following five years. Moreover, the ARIMA model was discovered as a better suited for forecasting Malaysia's short-term birth rate. Thus, for prediction, the ARIMA (0,0,2) model was employed. The growth birth rate from 2001 to 2022 was computed and found to be -0.0444, indicating a continuous decline in Malaysia's fertility rate of 0.0444 per year over the next five years. Furthermore, the fertility rate would continue to drop, if the government did nothing to address the issues with conception. As a result, Malaysia's fertility rate is experiencing an unprecedented crisis.

Keywords: birth rate; Malaysia; ARIMA Model; ETS Model; forecasting

Introduction

Malaysia's population is diverse and multiethnic, and its dynamic socioeconomic developments are reflected in the significant demographic changes that occurred between 2001 and 2022. In the early 2000s, there were about 23.5 million people living in Malaysia. This number increased to over 32 million by 2018, indicating rapid population growth driven by rising birth rates and advancements in healthcare that lowered infant mortality and raised life expectancy.

Regional and ethnic disparities also had an impact on the fertility rate dynamics. In comparison with Chinese and Indian populations, who had previously embraced the norm of having a small family, Malays had generally higher fertility rates. However, by 2022, these disparities were diminishing because socioeconomic changes and modern family planning techniques had an impact on every community.

According to the background analysis, Malaysia is currently facing serious problems with fertility and aging. Severe fertility problems affect Malaysia's long-term economic growth in addition to that level of compensation. They affect each aspect of lives of individuals, thereby affecting society as a whole.

First and foremost, comprehension of the birth rate has significant consequences for economic growth, social policies, and infrastructure design. In addition, lawmakers may be able to anticipate and get ready for the gravity of the future circumstances with the help of birth rate modeling and prediction. Furthermore, Malaysia's birth rate has been significantly influenced by government policies. Therefore, researching the results of these policies may aid in guiding future legislation and population control measures.

Moreover, forecasting birth rates in Malaysia using time series models is crucial for socio-economic and policy planning. These models, such as ARIMA and exponential smoothing, are used to analyze the historical data patterns and predict future trends. Furthermore, they can help to identify the seasonal variations and long-term trends, by providing the insights into potential population growth or decline. This information can help policymakers make informed decisions about resource allocation, infrastructure development, and social programs to support a growing or shrinking population.

In addition, the accuracy of birth rate forecasts also contribute to economic planning, ensuring the labor market, housing, and public services are adequately prepared for demographic changes. Thus, the application of time series models in birth rate forecasting is vital for sustainable development and effective governance in Malaysia.

Apart from that, this study aims to determine which model is more beneficial for Malaysia's short-term birth rate forecasting. The ARIMA and ETS models' capacity for prediction are based on a dataset comprising historical data on the birth rate of the nation from 2001 to 2022.

Literature Review

Introduction

According to Federal Reserve Bank of St. Louis, total fertility rate refers to a woman's ability to have children during her lifetime, which is affected by various factors such as social, biological, structural, and behavioral aspects. Poverty and income inequality have played a significant role in the flattening of fertility rates in certain countries, such as Brazil and developing nations. The decline in fertility rates has been linked to the state of education today, with birth probabilities likely being concentrated among women with lower levels of education.

Fertility is closely linked to public health and welfare systems, and since the 1950s, the global population fertility rate trend has been examined. The average fertility rate of women in developed countries is steadily declining from three to two children, while developing countries have seen a decrease. Countries like Europe, North America, Japan, Korea, China, Thailand, Australia, New Zealand, Brazil, Colombia, and Chile have low fertility rates, while regions like Southeast Asia, South Asia, North Asia, North Africa, Central America, and South America have intermediate fertility rates (John.B & Dennis.H. 2022)

Malaysia experienced a gradual decline in fertility rates from 2001 to 2018, reflecting shifting societal and economic dynamics. From 2000 to 2018, Malaysia experienced fluctuations in live births, influenced by socio-economic and demographic factors. The country underwent a significant demographic transition marked by a gradual but discernible decline in its growth rate, with urbanization playing a pivotal role in altering family structures and influencing fertility preferences. The government's proactive stance on family planning, investments in education, and the shift towards a knowledge-based economy contributed to declining fertility rates (Nurfarahain,M.S. & Subramaniam,G. , 2014).

Malaysia's crude birth rate has been declining due to changes in societal, economic, and healthcare dynamics. The decline is attributed to factors such as improved healthcare infrastructure, decreased infant mortality rates, and cultural norms favoring larger families. Factors contributing to this decline include increased access to family planning services, women's empowerment, and changing attitudes towards family size. Government policies promoting reproductive health and economic development likely also played a role in shaping these trends. ARIMA models, widely used in modeling and forecasting population variables since 1970, have been used to calculate future fertility rates, helping to inform decisions on birth control, family planning, and welfare policies. ARIMA and exponential smoothing models are recommended for predicting fertility rates, as they are commonly used in time series forecasting (Gotmark, et al, 2020).

In Malaysia, the use of ARIMA and ETS methodologies is widely used for birth rate forecasting. ARIMA models are effective in capturing linear aspects and time-dependent structures in historical data, providing a comprehensive approach for medium to long-term forecasting. ETS models are known for their ability to handle data with inherent trend and seasonal components, making them responsive to recent shifts in birth rates. The comparative analyses show that ARIMA models excel in environments with minimal seasonal effects, while ETS models provide superior accuracy in significant seasonality and non-linear trends. Thus, the integrating ARIMA and ETS models is recommended for better predictions and strategic planning.

Factors Associated with The Declining Evolution of Fertility and Birth Rates in Malaysia

The decline in fertility and birth rates in Malaysia is attributed to factors such as a fragmented labor market, challenges in balancing work and family obligations, and lack of childcare options. Research on fertility and reproductive behaviors has primarily focused on economic determinants and the relationship between fertility and economic conditions. The study of births and fertility is particularly relevant in local contexts, where demographic trends, socioeconomic characteristics, and geographic location vary greatly. Issues related to fertility dynamics include aging, depopulation, poor accessibility, and social and economic marginalization.

The Significance of Geographic Levels in The Analysis of Trends and Patterns in Fertility

The study examines fertility rates in Malaysia from 2001 to 2022, focusing on the years leading up to and following the 2008 financial crisis. It reveals a shift in the heterogeneity of macro-scale fertility patterns due to the fragmentation and mixture of micro-scale demographic behaviors, highlighting the intricate interaction of micro- and macro-factors of change. The research highlights the importance of examining fertility rates variability.

Methodology

Box Jenkins Models

According to Box, et al. (2008), Box-Jenkins methodology, named after George Box and Gwilym Jenkins, is a systematic approach for identifying, fitting, and validating Autoregressive Integrated Moving Average (ARIMA) models for time series data. It consists of three stages: model identification, parameter estimation, and model checking. It is widely used for accurate forecasting and understanding temporal dynamics, as it is robust and effective at modeling a wide range of time series data.

ARIMA Model

ARIMA models are widely used for time series forecasting due to their ability to handle various data patterns, including trends and seasonality. They consist of three components: autoregression, differencing to achieve stationarity, and moving averages. The parameters are *p*, *d*, and *q*.

The general formula for an ARIMA (*p,d,q*) model is:

$$\Phi_p(B)(1 - B)^d y_t = \Theta_q(B)\epsilon_t \tag{1}$$

where:

y_t is the observed time series
B is the backshift operator
$\emptyset_P(B)$ is the autoregressive (AR) operator defined as $1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_P B^P$
$\theta_q(B)$ is the autoregressive (MA) operator defined as $1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$
$\in_t$ is white noise error term

ARIMA models are useful tools for forecasting and analyzing time series data, providing insights into finance, economics, and environmental science. They involve determining appropriate values, estimating parameters, and diagnosing the model's adequacy using autocorrelation function plots.

AIC, AICc and BIC

The Akaike Information Criterion (AIC) and Bayesian Information Criterion are statistical metrics used in time series analysis to balance model complexity and goodness of fit, preventing overfitting.

This formula is used to calculate the AIC:

$$AIC = 2k - 2ln(L) \tag{2}$$

where:

k is the number of parameters in the model
L is the maximum value of the likelihood function for the model
$2k$ is penalized for the number of parameters, encouraging simpler models

A better model is indicated by a lower AIC value.

AICc, or corrected AIC, addresses bias in smaller datasets by adding a term to account for sample size, making it more trustworthy.

The AICc formula is:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{3}$$

where:

n is the number of sample in the dataset

AICc is a widely used method in scientific domains like ecology, economics, and engineering, where small sample sizes make AICc ineffective. It balances complexity and fit, enhancing model selection and evaluation. AICc is particularly useful in real-world scenarios where predictive accuracy and simplicity are crucial. In contrast, BIC has a larger penalty for models with more parameters.

The BIC is calculated using the formula:

$$BIC = ln(n)k - 2ln(L) \tag{4}$$

where:

n is the number of observations in the dataset
$\ln(n)\, k$ is a sample-size-based penalty that increases more rapidly with increasing n than the AIC's penalty does

The BIC becomes stricter in choosing models, favoring simpler ones as sample sizes increase. These standards are crucial in time series analysis, comparing models in ARIMA (AutoRegressive Integrated Moving Average) modeling. AIC tends to favor slightly overfitted models, while BIC is more conservative, favoring slightly under-fitted ones. Both AIC and BIC are essential tools for selecting models that fit well with fresh data.

Exponential Smoothing Model & Single Exponential Smoothing

According to Rob J. Hyndman and George Athanasopoulos, ETS models, also known as Exponential Smoothing State Space Models, are a popular method for forecasting time series data, particularly for data with patterns like trends and seasonality. It have three components: error (E), trend (T), and seasonal (S). The model can be additive or multiplicative, and is typically denoted as ETS(E,T,S).

The general formula for an ETS model,

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \tag{5}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t \tag{6}$$

$$b_t = b_{t-1} + \beta\epsilon_t \quad (7)$$

$$s_t = s_{t-m} + \gamma\epsilon_t \quad (8)$$

where:

y_t is the observed value at time t

l_t is the level component

b_t is the trend component

S_t is the seasonal component

$\in_t$ is the error term at time t

α, β, γ are smoothing parameters

m is the length of the seasonal cycle

ETS models are versatile tools used in various fields like business, economics, and inventory management to smooth past data and forecast future values, ensuring accurate and reliable predictions based on recent observations.

Single Exponential Smoothing, is used to forecast time series data by giving more weight to recent observations while gradually decreasing the weight of older data points. The formula is:

$$F_t = \alpha A_{t-1} + (1-\alpha)F_{t-1} \quad (9)$$

where:

F_t is the forecast for tim period t

A_{t-1} is rhe actual value at time period $t-1$

F_{t-1} is the forecast for time period $t-1$

α is the smoothing constant $(0 < \alpha < 1)$

This method is particularly useful for short term forecasting when data lacks significant trends or seasonal patterns, as it smooths out the random fluctuations and highlights the overall directions of the data.

Root Mean Square Error

Root Mean Square Error (RMSE) is a statistical metric used to evaluate the accuracy of a prediction model by comparing predicted values with observed data.

The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{t=1}^{n}(y_t - \hat{y}_t)^2} \quad (10)$$

where:

n is the number of observations

y_t is the actual values

$\hat{y}_t$ is the predicted values

RMSE, a sensitivity metric, is useful for evaluating model performance in finance, meteorology, and engineering. It's sensitive to large errors, making it useful for accurate predictions, but can be disadvantageous in outlier-heavy datasets.

Stationary Test: ADF Test

The Augmented Dickey-Fuller (ADF) test is a statistical method used to determine if a time series is stationary, ensuring constant mean and variance over time, and addressing higher-order autocorrelation.

The test requires estimating the following regression model:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \Sigma_{i=1}^{p} \delta_i \Delta y_{t-i} + \epsilon_t \quad (11)$$

where:

y_t is the time series being tested
Δy_t is the first difference of y_t
t is the time trend (optional)
α is a constant
β is the coefficient on the time trend
γ is the coefficient of the lagged level of th series
p is the number of lagged difference terms included (determined by criteria such as AIC or BIC)
δ_i is the coefficients on the lagged differences
$\in_t$ is the white noise error term

The ADF test is a crucial tool in time series analysis, determining the null hypothesis (H_0) and alternative hypothesis (H_1). If the p-value of test statistic is less than the critical value of the Dickey-Fuller distribution, the null hypothesis is rejected, indicating the series is stationary, and if it is greater, it fails to reject it.

Results and discussion

Data and Indicators

The study examines the use of crude birth rate (CBR) and total fertility rate (TFR) metrics at different spatial scales. TFR measures the typical number of children a woman would have if subjected to specific fertility rates. CBR, calculated by dividing the annual live birth rate by midyear population, is crude and focuses on local fertility. Both metrics were calculated from 2001 to 2022.

REGION	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
JOHOR	21.7	20.6	19.7	19.0	18.4	18.1	18.1	17.8	17.5	16.7	17.1	17.5	16.5	17.0	16.8	16.3	16.5	16.3	16.0	14.2	12.4	12.0
KEDAH	22.3	21.1	20.1	19.6	18.8	18.3	18.1	18.6	18.7	18.0	18.2	18.2	17.2	17.7	17.5	17.0	17.1	16.5	15.9	15.9	15.4	14.2
KELANTAN	27.5	26.3	24.8	24.0	22.9	22.3	22.1	22.5	22.7	22.6	23.3	22.4	21.6	22.8	22.2	21.4	21.4	20.6	19.6	20.2	19.2	18.5
MELAKA	21.5	20.2	19.8	18.8	18.1	17.1	17.4	17.2	16.7	15.8	16.8	16.9	16.0	16.7	16.4	16.0	15.8	15.6	15.2	13.9	13.1	13.0
PAHANG	20.0	19.6	18.8	18.3	17.7	17.2	17.5	17.4	17.1	16.5	16.9	16.9	16.3	17.3	16.9	16.4	16.2	16.2	15.5	14.3	13.9	13.2
NEGERI SEMBILAN	21.0	20.0	19.2	18.8	18.3	17.5	17.3	17.6	17.6	17.3	17.7	17.7	17.2	18.1	17.5	17.0	16.7	16.2	15.5	15.9	15.3	14.6
PERAK	20.4	19.6	18.2	17.6	16.6	16.2	15.9	15.9	15.8	15.1	15.6	15.5	14.5	14.9	14.7	14.4	14.0	13.5	13.2	12.7	12.2	11.7
PERLIS	20.6	19.8	18.8	18.1	17.7	17.3	17.4	18.1	18.2	17.4	17.5	17.3	16.7	18.4	17.4	17.5	17.6	17.2	16.0	14.4	14.2	12.9
PULAU PINANG	18.2	17.9	16.6	16.0	15.4	14.8	14.9	15.0	14.7	13.4	14.0	14.7	12.9	13.6	13.1	12.7	12.3	11.7	11.4	11.3	10.7	10.2
SARAWAK	21.8	20.8	20.0	19.4	18.9	18.4	18.0	18.0	17.9	16.7	17.2	16.8	15.3	15.4	14.7	13.9	13.6	13.1	13.1	14.4	12.7	12.5
SELANGOR	20.3	20.2	19.6	19.8	19.1	19.0	18.9	19.0	19.0	17.9	18.5	18.9	17.5	18.0	17.5	16.6	16.3	15.8	15.2	13.4	12.8	11.9
TERENGGANU	26.2	24.8	23.4	23.3	22.4	22.0	22.1	22.8	23.0	23.0	23.0	23.5	22.9	23.7	23.8	23.3	23.1	23.0	21.8	23.6	22.5	21.3
KUALA LUMPUR	18.8	17.6	16.8	16.3	15.2	15.3	15.6	15.5	15.7	14.6	15.1	15.9	14.7	15.6	14.5	14.4	13.8	13.5	13.4	11.7	10.7	10.1
LABUAN	19.3	18.8	20.1	19.0	18.7	19.4	19.0	19.9	19.2	18.5	19.4	19.6	19.1	19.9	20.5	18.1	17.3	17.1	16.9	18.0	15.9	14.2

Table 1: CBR by administrative regions of Malaysia from 2001 – 2022

Data Processing

The Malaysian fertility rate's time series is non-stationary, with no trends or seasonality. This makes it difficult to model the data, as the mean, variance, and correlation structure may change over time. To obtain a stationary time series, the data was transformed into a growth rate, a growth factor, and the raw birth rate data. This process standardizes the series mean and reduces variation, making it suitable for statistical modeling techniques like ARIMA.

Year	Live Birth	Crude Birth Rate	Year Difference	Value Difference	Growth Rate	Growth Factor
2001	505479	21.0	-	-	-	-
2002	494538	20.2	1	-0.8	-0.0380952	0.9619048
2003	481399	19.2	1	-1.0	-0.0495050	0.9504950
2004	481800	18.9	1	-0.3	-0.0156250	0.9843750
2005	474473	18.2	1	-0.7	-0.0370370	0.9629630
2006	472698	17.8	1	-0.4	-0.0219780	0.9780220
2007	479647	17.7	1	-0.1	-0.0056180	0.9943820
2008	493203	17.9	1	0.2	0.0112994	1.0112994
2009	501644	17.9	1	0.0	0.0000000	1.0000000
2010	491239	17.2	1	-0.7	-0.0391061	0.9608939
2011	511594	17.6	1	0.4	0.0232558	1.0232558
2012	526012	17.8	1	0.2	0.0113636	1.0113636
2013	503914	16.7	1	-1.1	-0.0617978	0.9382022
2014	528612	17.2	1	0.5	0.0299401	1.0299401
2015	521136	16.7	1	-0.5	-0.0290698	0.9709302
2016	508203	16.1	1	-0.6	-0.0359281	0.9640719
2017	508685	15.9	1	-0.2	-0.0124224	0.9875776
2018	501945	15.5	1	-0.4	-0.0251572	0.9748428
2019	489863	15.1	1	-0.4	-0.0258064	0.9741936
2020	471504	14.5	1	-0.6	-0.0397351	0.9602649
2021	439744	13.5	1	-1.0	-0.0689655	0.9310345
2022	423124	12.9	1	-0.6	-0.0444444	0.9555556

Table 2: Crude Birth Rate Values and Calculation Data related to Crude Birth Rate Values for each year in Malaysia since 2001.

Results After Computing the Growth Rate and the Growth Factor

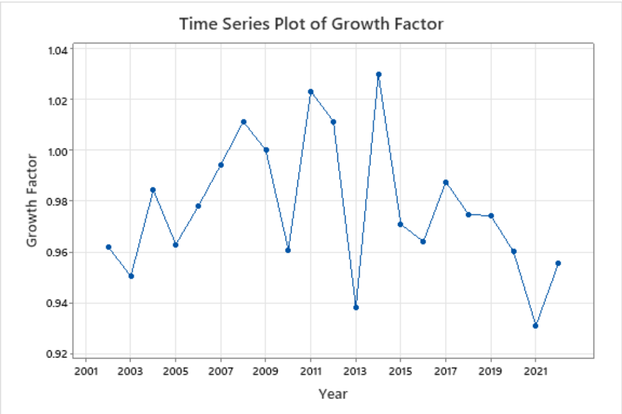


Figure 1: Time Series Plot for Growth Factor of Crude Birth Rate from 2001 to 2022

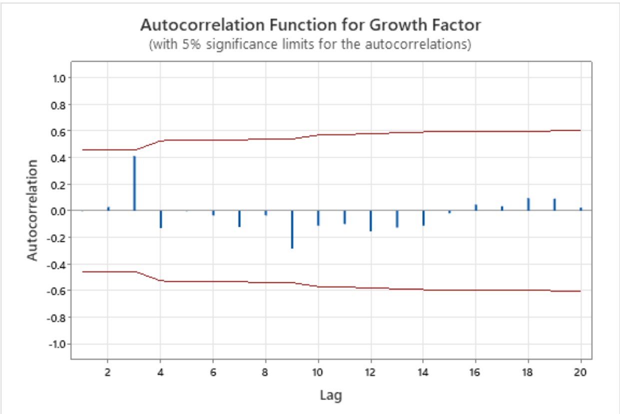


Figure 2: Autocorrelation Function for Growth Factor of Crude Birth Rate

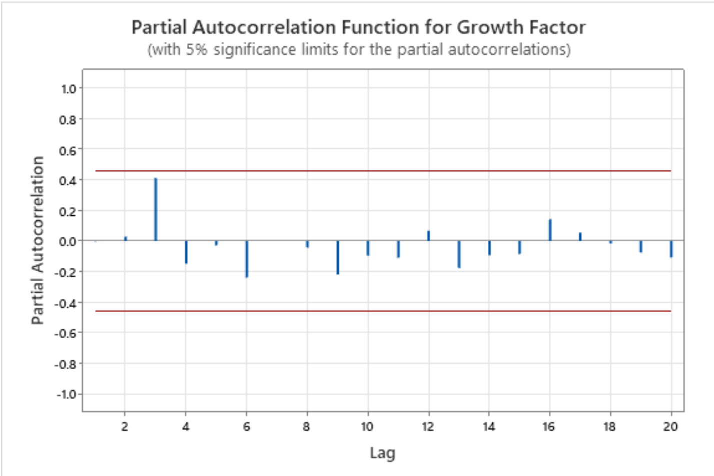


Figure 3: Partial Autocorrelation Function for Growth Factor of Crude Birth

Table 4.2 above shows Malaysia's crude birth rate figures and the corresponding calculation data for each year since 2001. In order to assist with data visualization and comprehension, Figures 1, 2, and 3 show the evolution of Malaysia's birth rate growth factor over time.

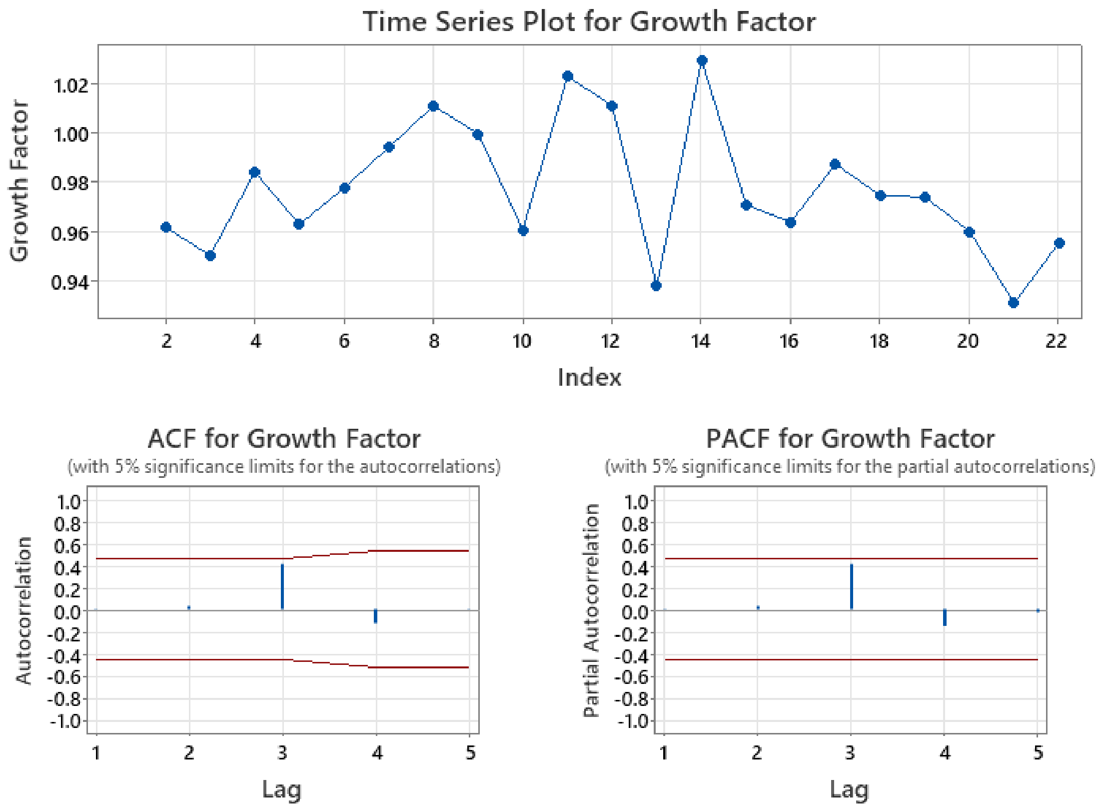


Figure 4: Time Series Plot for Growth Factor by Using Augmented Dicker Fuller Test

The p-value of 0.919, which is greater than the 5% significance level, was determined using the Augmented Dicker Fuller Test. As a result, it was concluded that the time series was non-stationary, and the null hypothesis was accepted.

Results After Computing the Growth Rate and the Growth Factor

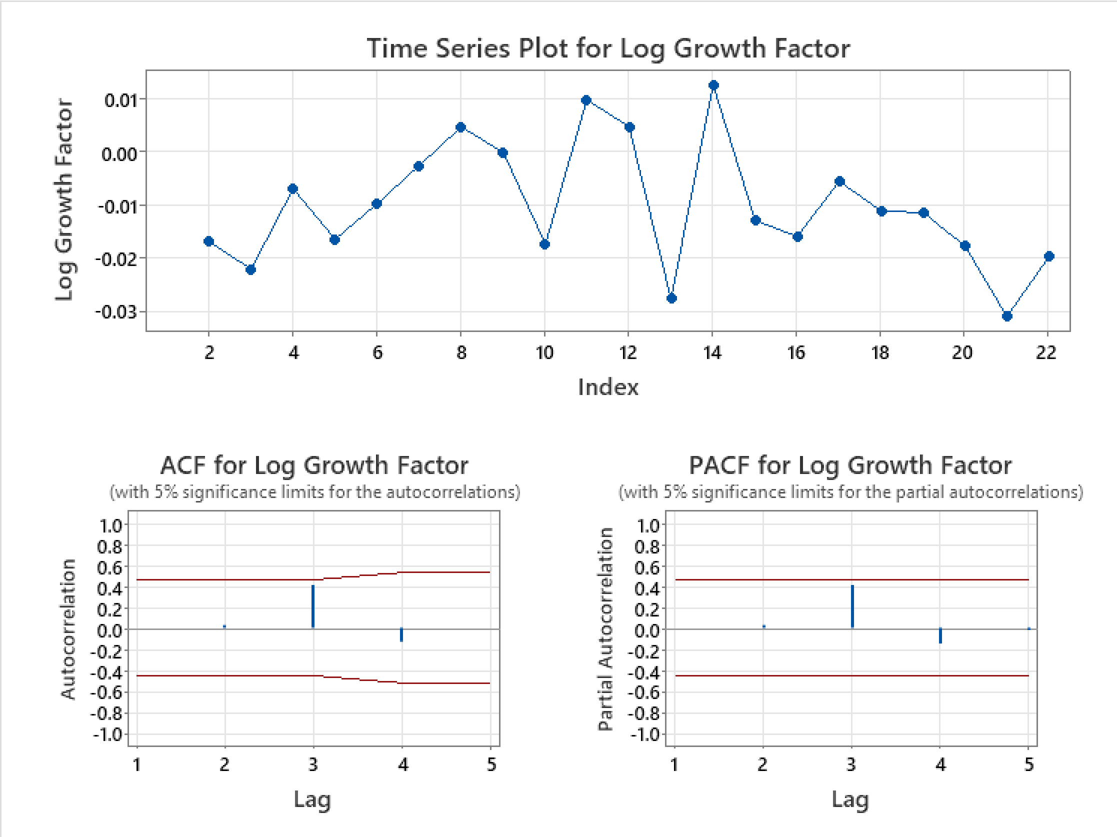
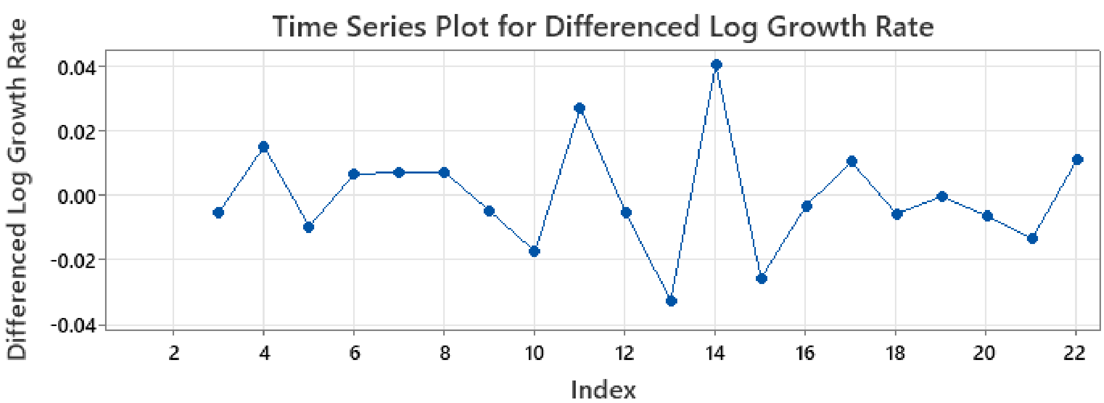


Figure 5: Time Series Plot for Log-Growth Factor by Using Augmented Dickey Fuller Test

The improvement of the variance stabilization of a time series is the aim of the logarithm transformation applied to the growth factor. While a time series shows exponential growth or decay, the logarithm of the data is particularly instructive in helping to linearize the data. The p-value, which indicated significance at the 5% level, was found to be 0.921 using the Augmented Dickey Fuller Test. It consequently becomes unable to conclude that the time series was non-stationarity and to rule out the null hypothesis. The time series remained non-stationary because the logarithm transformation was unable to further normalize the growth factor data.

Results After Differencing



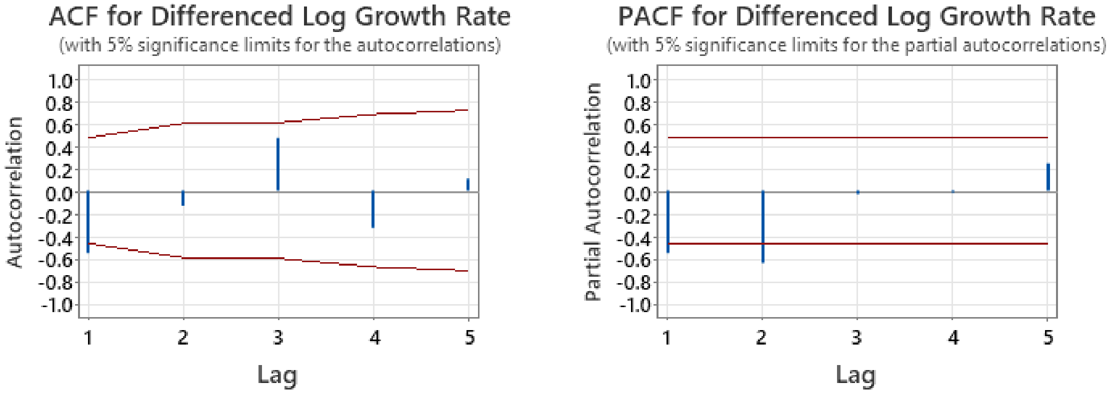


Figure 6: Time Series Plot for Differenced Log-Growth Factor by Using Augmented Dickey Fuller Test

By subtracting the previous observation from the current observation, the data was separated in a desire to make the time series statistically stable. The p-value was determined using the Augmented Dickey Fuller Test with a significance level of less than 5%. The purpose is to identify whether the time series was stationary as the result of the rejection of the null hypothesis. The best data to use in our investigation is the crude birth rate data's log transformed growth factor after one differencing, as the transformed data time series was found to be stationary by the mentioned earlier Augmented Dickey Fuller Test.

Data Partitioning

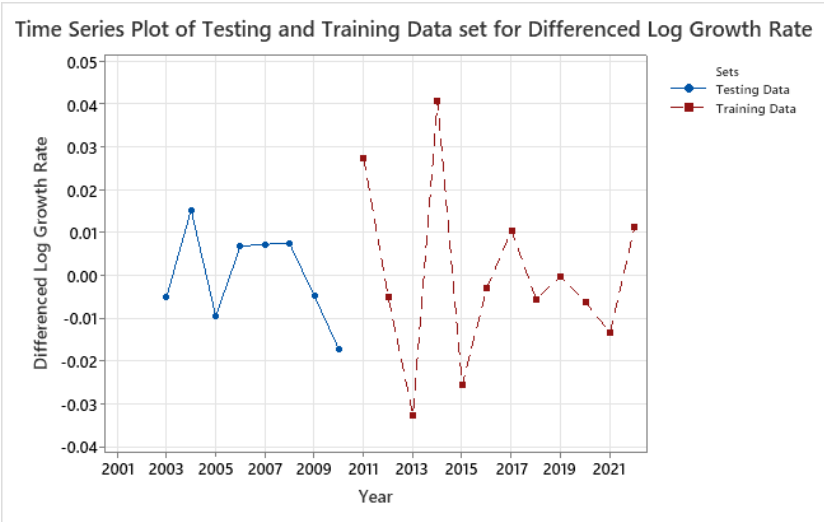


Figure 7: Time Series Plot of Testing and Training set for Differenced Log-Growth Factor

The transformed data was divided into two sets: the training set was used to fit the data, and the testing set was used to ensure the accuracy of the model's forecasts. The data used in the testing set was from 2011 to 2022, and the training set included data from 2001 to 2010.

ARIMA Model

Model Description

Minitab's Forecast with Best ARIMA Models was utilized for forecasting and ARIMA modeling, generating the ARIMA(0,0,2) model with 0,0, and 2 as parameters.

Model Selection

Model (d = 0)	LogLikelihood	AICc	AIC	BIC
p = 0, q = 2*	62.2793	-113.892	-116.559	-112.576
p = 2, q = 0	62.2087	-113.751	-116.417	-112.434
p = 2, q = 1	62.1570	-113.647	-116.314	-112.331
p = 0, q = 1	59.1479	-110.796	-112.296	-109.309
p = 1, q = 1	59.7605	-108.854	-111.521	-107.538
p = 1, q = 0	57.2396	-106.979	-108.479	-105.492
p = 2, q = 2	61.9173	-105.373	-111.835	-105.860
p = 1, q = 2	59.3486	-104.412	-108.697	-103.719

** Best model with minimum AICc. Output for the best model follows.*

Table 3: ARIMA Model Selection

According to the results that shown in Table 4.5, ARIMA(0,0,2) model have the lowest AICc values which is -113.892. So, ARIMA(0,0,2) model would be the suitable model for using to forecast the birth rates in study.

Accuracy Test

The ARIMA model was utilized to predict years in the testing set, with Minitab's Forecast with Best ARIMA Models being used to assess accuracy.

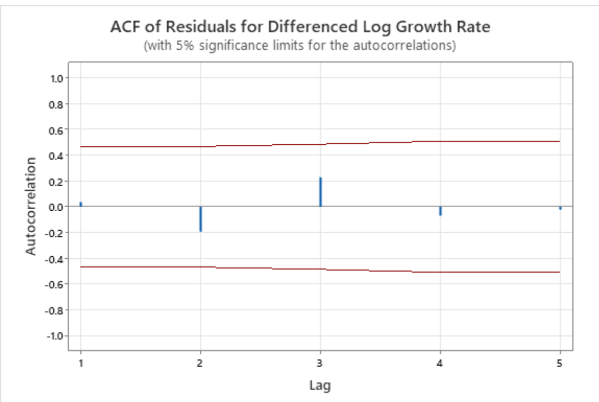
Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
MA 1	1.267	0.144	8.77	0.000
MA 2	-0.966	0.148	-6.54	0.000
Constant	-0.00062	0.00141	-0.44	0.664
Mean	-0.00062	0.00141		

Model Summary

DF	SS	MS	MSD	AICc	AIC	BIC
17	0.0018120	0.0001066	0.0000906	-113.892	-116.559	-112.576

MS = variance of the white noise series



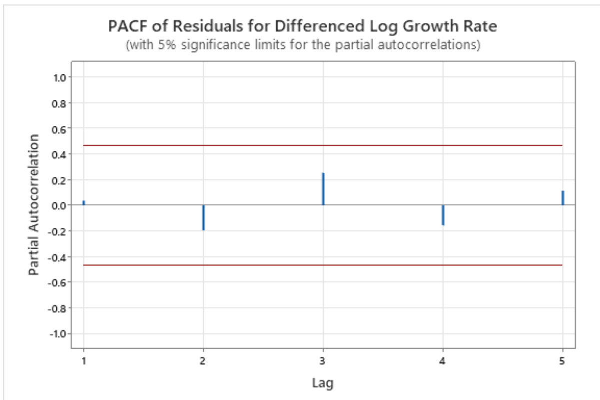


Figure 8: Time Series Plot of Testing and Training set for Differenced Log-Growth Factor

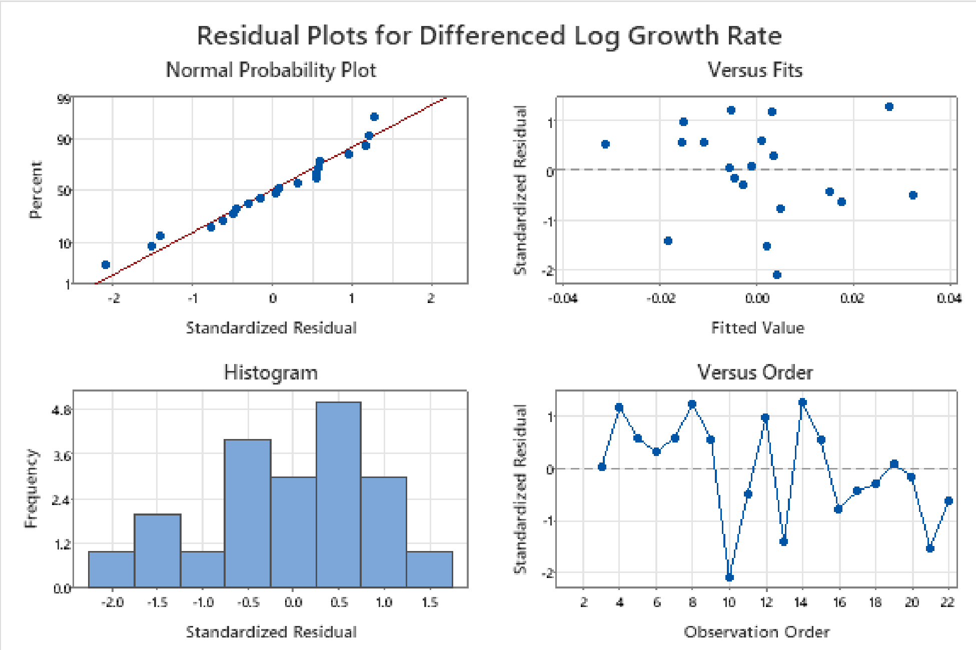


Figure 9: Residual Plots for Differenced Log-Growth Factor

From the figure provided above a comprehensive summary of the model performance and parameter estimates was shown. The Model Summary table shows that the model has 17 degrees of freedom (DF), with a Sum of Squares (SS) of 0.0018120, Mean Squares (MS) of 0.0001066, and Mean Squared Deviation (MSD) of 0.0000906. The model selection criteria values are AICc = -113.892, AIC =-116.559, and BIC = -112.576, indicating a good fit as lower values suggest as a better models.

The Final Estimates of Parameters table shows the parameter estimates for the model. The first Moving Average (MA) term has a coefficient of 1.267, with a standard error (SE) of 0.144, a T-value of 8.77, and a highly significant p-value of 0.000, indicating its strong significance. The second MA term has a coefficient of -0.966, SE of 0.148, T-value of -6.54, and a p-value of 0.000, also showing a strong significance. The constant term has a coefficient of -0.00062, SE of 0.00141, T-value of -0.44, and a p-value of 0.664, indicating it is not statistically significant. Furthermore, the mean is -0.00062 with an SE of 0.00141.

The Modified Box-Pierce (Ljung-Box) Chi-Square Statistic table evaluates the model adequacy through residuals autocorrelation. At a lag of 12, the Chi-Square value is 6.19 with 9 degrees of freedom and a p-value of 0.721. The high p-value suggests that there is no significant autocorrelation in the residuals, indicating the model adequacy in capturing the data patterns.

ETS Model

Model Description

The Single Exponential Smoothing method was used to automatically select the appropriate taxonomy and model constants. The taxonomy describes each model in three dimensions: seasonality, trend, and error. The paper states that those three dimensions can be classified as "additive," "multiplicative," or "none." Under these circumstances, the ETS (A,N,N) model—simple exponential smoothing with additive errors—was automatically selected.

Accuracy Test

The ETS model was utilized to predict years in the testing set, with Minitab calculating accuracy using the Single Exponential Smoothing method.

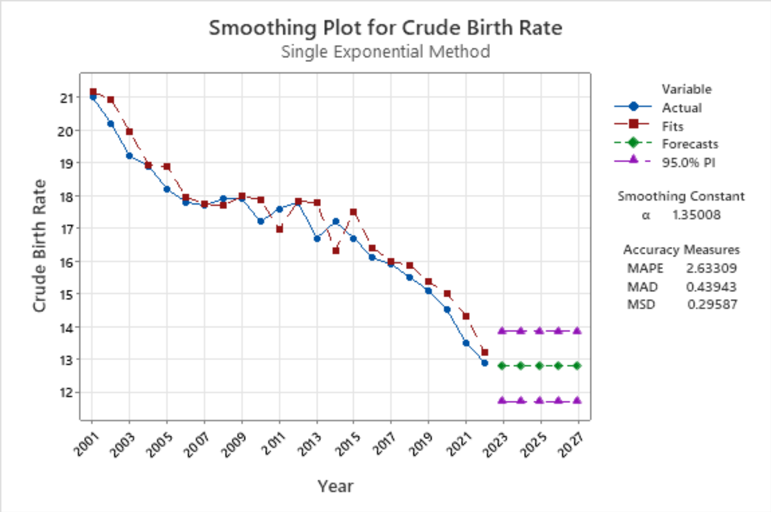


Figure 10: Smoothing Plot for Crude Birth Rate

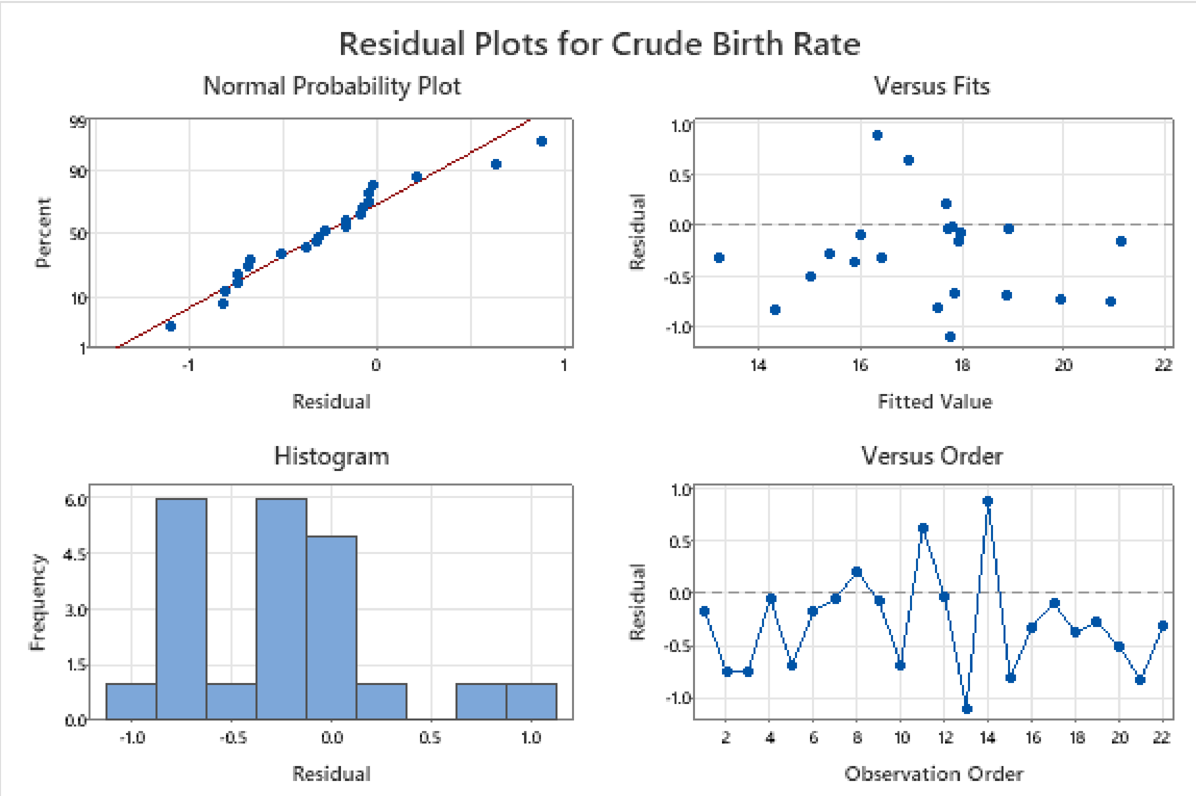


Figure 11: Residual Plots for Crude Birth Rate

For Figure 10 show a smoothing plot that illustrating the actual crude birth rates from 2001 to 2021 and the fitted values based on the Single Exponential Smoothing method. The actual values are marked in blue, the fitted values in red, and the forecasts for 2023 to 2027 in green, with purple arrows indicating the 95% prediction intervals. The plot indicates a steady decline in the crude birth rate from around 21 in 2001 to approximately 13 in 2021, followed by a stabilization in the forecasts.

For Figure 11 displays four residual plots for the crude birth rate. The Normal Probability Plot shows that the residuals are approximately normally distributed, indicated by the close alignment with the straight line. The Versus Fits plot shows the residuals scattered randomly around zero, suggesting no apparent pattern and that the model's fit is reasonable. Other than that, the Histogram of residuals indicates a roughly normal distribution, while the Versus Order plot shows the residuals over the order of observations, suggesting no significant autocorrelation, as the residuals appear randomly dispersed.

Overall, the analysis demonstrates a consistent downward trend in the crude birth rate with reliable forecasting, validated by residual analysis that supports the adequacy of the single exponential smoothing model used.

Five year Forecasts Using ARIMA(0,0,2) Model

The accuracy tests on both models show ARIMA(0,0,2) is more precise for forecasting Malaysia's crude birth rates over the next five years, using the same training set.

Forecasts

			80% Limits		95% Limits		
Time Period	Forecast	SE Forecast	Lower	Upper	Lower	Upper	Actual
23	-0.0006244	0.0194254	-0.0255226	0.0242737	-0.0387059	0.0374571	
24	-0.0006244	0.0194254	-0.0255226	0.0242737	-0.0387059	0.0374571	
25	-0.0006244	0.0194254	-0.0255226	0.0242737	-0.0387059	0.0374571	
26	-0.0006244	0.0194254	-0.0255226	0.0242737	-0.0387059	0.0374571	
27	-0.0006244	0.0194254	-0.0255226	0.0242737	-0.0387059	0.0374571	

Table 4: A five year forecast of the birth rate in Malaysia using ARIMA(0,0,2) Model.

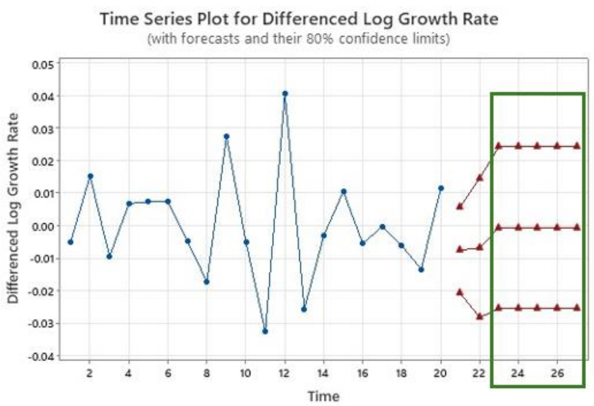


Figure 12: 5 Year Forecast in 80% confidence limits using ARIMA(0,0,2) Model

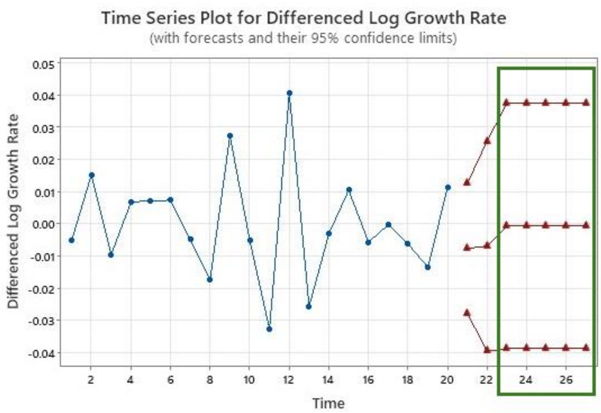


Figure 13: 5 Year Forecast in 95% confidence limits using ARIMA(0,0,2) Model

Results

The ARIMA(0,0,2) model predicts a -0.0006244 growth factor difference between 2023 and 2027, resulting in a 0.04444 annual drop in Malaysia's birth rate. This indicates a steady decline in fertility rates for the next five years, with a growth factor of 0.9556 for 2022.

Discussion

The study compared the ARIMA and ETS models for forecasting Malaysia's crude birth rate. The ARIMA model showed comparable trends in time series plots, with the ETS model predicting a steady increase between 2011 and 2022. The ARIMA model showed an initial increasing trend before stabilizing over the next five years. Both models projected Malaysia's birth rate to increase at a constant rate from 2023 to 2027. The ARIMA model performed better in terms of AIC and BIC values, but the ETS model had a lower ME value and higher residuals p-value. The ARIMA model may not be the most accurate for predicting future birth rates.

Conclusion

The most appropriate information to use through data processing is the log-transformed growth factor of the original birth rate data after one differencing, as its p-value is significantly smaller than the significance level at 5%, indicating that stationary nature of this time series. The ETS model and the ARIMA model were used on the training set of birth rate data, respectively. The accuracy test results indicate that the ARIMA(0,0,2) model is a better fit for predicting Malaysia's future birth rate. The ARIMA(0,0,2) model's forecasting indicates that in the five years starting in 2022, the difference in the growth factor from year to year would be 0.9556. In conclusion, during the next five years, Malaysia's birth rate will fall at a steady pace of 0.04444 per year. As a result, Malaysia's crude birth rate is currently experiencing an unparalleled crisis.

Acknowledgement

My sincere gratitude goes out to everyone who helped make this study on Time Series Forecasting of Crude Birth Rate in Malaysia a success. I am particularly thankful to my mentors and academic advisors for their invaluable advice, insights, and support during the course of the study. I am also grateful to the statistical and data analysis experts whose knowledge was crucial for assisting me refine my methodologies. Lastly, I would like to embrace my appreciation to the institutions and organizations that provided the data and resources that were necessary to complete this paper.

References

[1] Alkema, L., Raftery, A.E., Gerland, P., Clark, S.J., Pelletier, F., Buettner, T. and Heilig, G.K. (2011). Probabilistic Projections of the Total Fertility Rate for All Countries. *Demography* 48: 815-839.

[2] Anuj, K. and Bhupendra, M. (2023). Comparative Analysis of the Time Series ARIMA Model over the AR, MA, and Holt's Models for Future Prediction of Total Fertility Rate in India. *The Pharma Innovation Journal*, vol. 12, no. 7, pp. 912-920.

[3] Billari, F. (2015). Integrating macro-and micro-level approaches in the explanation of population change. *Population Studies,* 69(sup1), S11–S20.

[4] Billari, F., and Kohler, H.-P. (2004). Patterns of low and lowest-low fertility in Europe. *Population Studies,* 58(2), 161–176.

[5] Public Sector Open Data, Department of Statistics Malaysia. (2024). Annual Live Births. https://open.dosm.gov.my/data-catalogue/births_annual

[6] Macrotrends LLC. (2010). Malaysian Population Growth Rate From 1950 to 2024. https://www.macrotrends.net/global-metrics/countries/MYS/malaysia/population-growth-rate

[7] Qingfeng, W. and Xu, S. 2016. The Role of Socio-political and Economic Factors in Fertility Decline: A Cross-country Analysis. *World Development* 87:360-370.

[8] Wei, Y. et al. (2019). Predicting Population Age Structures of China, India, and Vietnam by 2030 based on Compositional Data. *PLoS One*, vol. 14, no. 4:e0212772.

[9] Yifan,W. (2023). Analysis of China's Birth Rate Prediction Based on Time Series. *Department of Statistics, The University of British Columbia,* V6T 1Z4, Vancouver, Canada.

[10] Yongjie, C. (2023). Model Analysis on the Birth Rate in China. *Department of Statistics, University of California, Davis One Shields Avenue, Davis,* CA 95616.