Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/59/20231010

Modeling and Prediction of Birth Rate in China

Yunhan Zhang[1,a,*]

[1]*Department of Economics, University of California, Davis, One Shields Avenue, Davis*
a. unzhang@ucdavis.edu
**corresponding author*

Abstract: Concerns about a persistent reduction are growing against the backdrop of large changes in China's birth rate during the last few decades. This paper explores this tendency via the lens of time series analysis, using birth rate records from 1964 through 2021 (particularly ignoring 1960-1963 due to Great Leap Forward distortions). The ARIMA and ETS models were extensively studied in our hunt for the best accurate forecasting device. The ARIMA (0,0,1) model was considered to be preferred based on comparison measures. The primary goal of this model was to predict the trend of China's fertility rates over the following five years. The ARIMA an1``d ETS models were rigorously applied to a selected training set after initial adjustments to ensure data stationarity, followed by an evaluation of their accuracy. Our findings, which are backed by the ARIMA model, imply a disturbing trend: a 0.117 percent annual fall in China's birth rate from 2022 to 2026. This suggests that a national fertility crisis is on the horizon. As a first step, we advise looking at the various socioeconomic reasons that may be driving this trend, as well as evaluating policy actions that could serve as potential cures.

Keywords: Birth rate, Time series analysis, ARIMA model, ETS model, Forecasting

1. Introduction

The study's significance is that the drop in the birth rate is more than just a demographic statistic. It has far-reaching repercussions for the nation's socioeconomic fabric. External reasons, such as the Covid-19 pandemic, cannot be overlooked, but intrinsic factors, specifically low fertility intention and postponement of reproduction due to growing living costs, have been the key drivers to this reduction. Addressing the dropping birth rate is no more a matter of policy preference; it has become a pressing issue for both the government and society as a whole.

Our research strives to provide foresight into this subject by recognizing the overarching relevance of the birth rate as a demographic indicator and its consequential impact on social and economic planning. To that purpose, we intend to apply the ARIMA (Autoregressive Integrated Moving Average) and ETS (Error, Trend, Seasonality) models on a data set spanning the years 1960 to 2021. This study is methodologically constructed to partition these data into dedicated training and testing sets, assuring the robustness of model fitting as well as the rigor of their prediction accuracy evaluation.

By embarking on this research endeavor, we hope to not only illuminate the patterns and potential trajectories of China's birth rate, but also provide policymakers and stakeholders with data-driven insights to assist in their strategic decision-making.

DOI: 10.54254/2754-1169/59/20231010

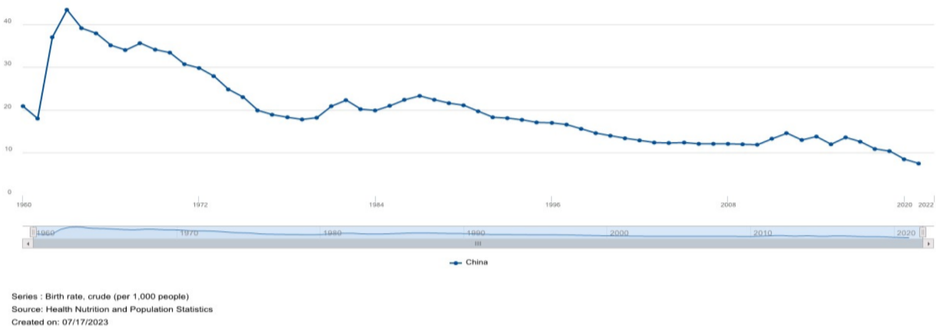

Figure 1: Patterns of China's Birth Rate

This paper begins with an exploration of China's declining fertility rates and the urgent need for precise forecasting models. In our methodology, we outline data partitioning and the adoption of the ARIMA and ETS analytical models, followed by a results section detailing our findings for the 2022-2026 period. We subsequently discuss the relative strengths and limitations of our study, especially concerning data selection and model choice. The paper concludes by synthesizing our insights on China's future demographic trends and suggests avenues for future research, emphasizing potential socio-economic influences on birth rates and potential policy interventions (Figure 1).

2. Motivation

2.1. Societal Implications of Birth Rate Dynamics

The birth rate, as an important demographic statistic, has far-reaching consequences that affect many facets of society. Its consequences extend beyond population statistics to include social policies, infrastructure planning, and the larger realm of economic development. The demographic distribution, which is influenced by the birth rate, has the ability to impact key socioeconomic pillars such as:

Healthcare: Age distribution affects healthcare needs, facilities, and policy strategies.

Education: The number of births determines the upcoming demands on educational facilities, from preschools to universities.

Social Security: An aging population invariably exerts pressure on the nation's social security system.

Labor Market: The labor force, its size, and its quality are intrinsically linked to birth rate dynamics.

2.2. The Imminent Threat of an Aging Population:

Every country has several obstacles as its population ages. Birth rate predictive modeling is an important tool for policymakers to get insight into anticipated demographic transitions. These forecasts enable proactive planning, allowing for timely and effective preparation to address the possible issues provided by an aging population.imely and effective preparedness to alleviate the possible challenges of an aging population [1].

2.3. Impact Analysis of Government Population Policies

Throughout history, China's government policies have significantly molded its demographic tapestry. Its birth rate trajectory has been affected in part by defining policies such as:

01. One-Child Policy (1979-2015): Initially implemented as a population control strategy, the socio-demographic consequences have been severe.

DOI: 10.54254/2754-1169/59/20231010

02. Two-Child Policy (2016-2021): A reform with the ability to alleviate the impacts of its predecessor.

03. Three-Child Policy (as of 2021): A government policy recognizing the need for demographic rejuvenation.

A thorough examination of the consequences of these policies provides not only a historical perspective, but also a roadmap for future legislative initiatives and creative population management measures.

3. Literature Review

For more than a decade, academic debate has focused on China's dropping birth rate and the related socio-demographic concerns.

3.1. The Historical Background of China's Birth Rate

The purpose of China's extended family planning program was to keep the country's population within sustainable boundaries by reining in the country's soaring birth rate. Despite achieving its primary goal, this line of action unintentionally caused demographic issues. Unintended repercussions of these measures included a significant tilt in China's demographic distribution toward the old, as well as the concomitant socioeconomic burden of maintaining this geriatric population [2].

3.2. Birth Rate Modeling Fundamentals

The dynamic fluctuations in the birth rate highlight the need of predictive modeling as a fulcrum for policy design across the birth control, family planning, and welfare frameworks. Given the complex web of contributing variables, emerging countries' need for accurate models suited to demographic metrics such as birth rate becomes even more pressing.

3.3. Time Series Forecasting Models

As the two foundations of time series forecasting, the paper emphasized the relevance of exponential smoothing (ETS) and ARIMA models.

ARIMA (AutoRegressive Integral Moving Average): The ARIMA model uses the autocorrelations inherent in the dataset [3]. It is based on the idea that previous time series data is a treasure of information for future forecasts.

ETS (Exponential Smoothing): Unlike ARIMA, ETS models are based on generating weighted averages of antecedent observations, providing a different but equally robust forecasting strategy [4].

4. Data Description and Processing

4.1. Data Description

The data under study was procured from the Health Nutrition and Population Statistics databank. Access it here: https://databank.worldbank.org/source/health-nutrition-and-population-statistics?l=en#

The data represents the birth rate in China, quantified as births per 1,000 individuals. This time series spans over 57 years, starting from 1963 and concluding in 2021. Each observation corresponds to a discrete year.

A notable fluctuation occurred about 1962 during early data assessment. The historical context for this fluctuation is provided by the 'Great Leap Forward' (1958-1962) and the 'Three Years of Great Chinese Famine' (1959-1961). These subsequent events had a significant impact on China's

DOI: 10.54254/2754-1169/59/20231010

socioeconomic structure. During the height of the famine in 1960, the birth rate fell to 14 per thousand, a considerable fall from the previous years' rate of 20 per thousand. After 1962, as a result of these difficulties being alleviated, there was a modest recovery in the birth rate, while it remained vulnerable to lingering effects.

For the sake of stability and reliability, our analysis is focused on data from 1963 to 2021(Figure 2).

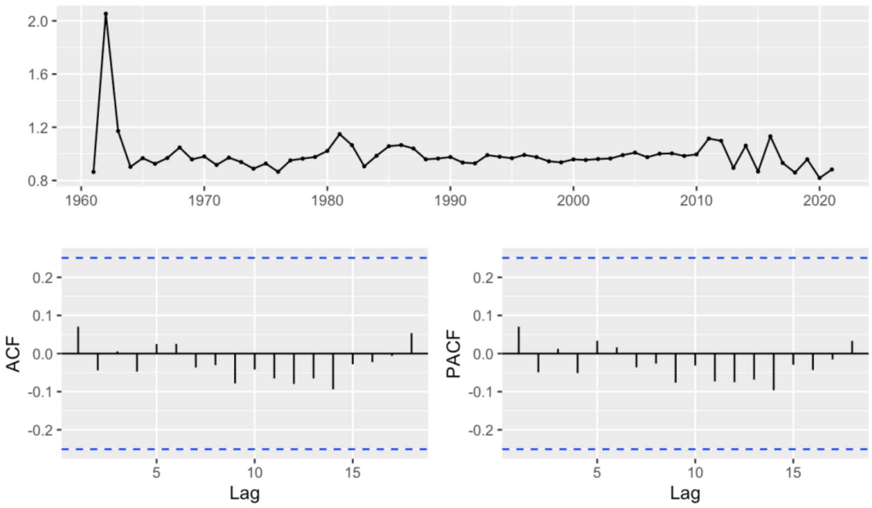

Figure 2: Time Series Plot for Birth Rate in China

4.2. Data Processing

A time series diagram, designated Figure 3, depicts a discernible downward trend in China's birth rate, indicating non-stationarity.

Understanding Stationarity Stationarity, a crucial characteristic of time series data, ensures that properties such as mean and variance remain constant over time. Numerous statistical instruments, ARIMA being the exemplar, are based on this assumption [5].

This test was conducted following each transformation. Initially, with a p-value of 0.3453 and then 0.359 after logarithmic transformation, it was determined that the time series was not stationary. After employing differencing, however, the p-value decreased to 0.01, indicating stationarity within our acceptable significance level of 5%.

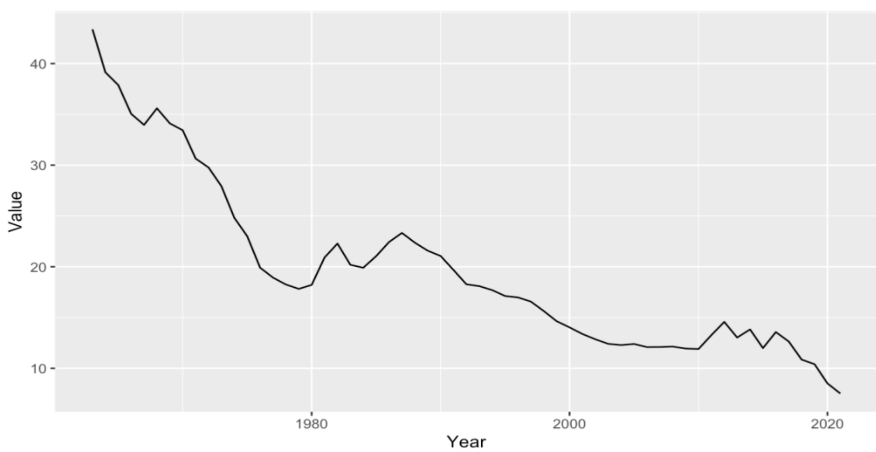

Figure 3: Time Series Plot for Birth Rate in China

DOI: 10.54254/2754-1169/59/20231010

From Figure 3, there is a clearly decreasing trend for the time series of the birth rate in China, so we could conclude that this series is non-stationary.

To further illustrate, stationarity is a critical property for time series data that many statistical modeling techniques require.

A stationary time series is one whose properties (like mean and variance) do not change over time. In other words, it doesn't exhibit trends or seasonality. Many statistical models, such as ARIMA, assume that the underlying data are stationary because these models are designed to predict the constant mean, variance, and autocorrelation structure in the data [6].

Non-stationary data, on the other hand, often contain trends or seasonal patterns. The mean, variance, or correlation structure may change over time, making the data harder to model. If these properties are changing, the patterns that the model learns may not apply to future periods [6].

To obtain the stationary time series, we decided to make some transformations for these data:

Firstly, transform data of average birth rate to the growth rate for each year;

Then, compute the growth factor to standardize the mean of the series by reducing the variation.

Then, take the logarithm of the growth factor to further standardize the data if needed.

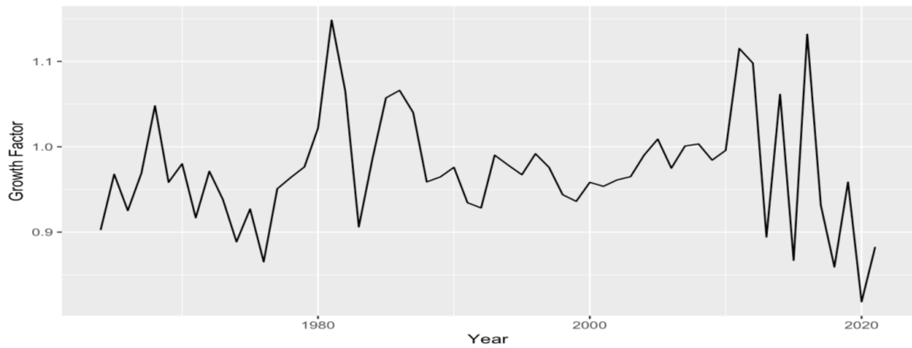

Figure 4: Time Series Plot for Growth Factor of Birth Rate in China

5. Methodology

5.1. Data partitioning

In order to enhance the resilience of our models and facilitate their subsequent evaluation, the altered dataset was partitioned into separate training and testing sets. The period spanning from 1964 to 2016 was designated for the purpose of training, whereas the data including the years 2017 to 2021 was set aside specifically for validation and model assessment.

5.2. ARIMA Model

5.2.1. Model Description

We employed the auto.arima() function in R to determine the optimal ARIMA model parameters. The function, grounded in the Hyndman-Khandakar algorithm [4], amalgamates unit root tests, AICc minimisation, and maximum likelihood estimation (MLE) to derive the best-fitting ARIMA structure. The selected model was ARIMA (0,0,1) with a zero mean, indicating a first order moving average model.

DOI: 10.54254/2754-1169/59/20231010

5.2.2. Accuracy test

After model construction, we projected birth rates for the years in our testing dataset. The juxtaposition of these forecasts with real data can be observed in Figure 5. Model accuracy was then gauged using R's accuracy() function, with the results detailed in Table 1.

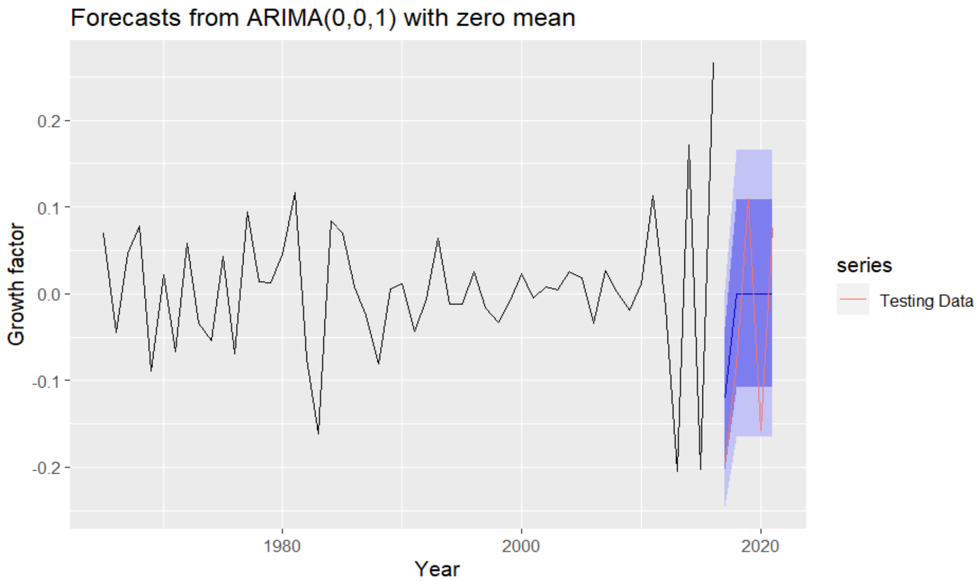

Figure 5: ARIMA (0,0,1) Model

Table 1: Summary for Accuracy Test of the ARIMA (0,0,1) Model

	ME	RMSE	MAE	MPE	MAPE	MASE
Training Set	0.00959091	0.06300125	0.04647275	97.99229	166.40961	0.5295333
Test Set	0.02557882	0.10455979	0.09956321	87.60218	87.60218	1.1344721

5.3. ETS Model

5.3.1. Model Description

The ets() function, an automatic model selection tool in R, was utilized to identify the optimal ETS model structure. This approach represents models across three facets: error, trend, and seasonality (ETS). Each element can exhibit additive, multiplicative, or no effect [3]. The model determined was ETS(A, N, N), signaling simple exponential smoothing with additive errors. Notably, the smoothing parameter α approximated to 0, suggesting a static series level [4].

5.3.2. Accuracy Test

Forecasts produced by the ETS model for the testing years are contrasted with real data in Figure 6. Model accuracy was evaluated using the accuracy() function, with findings presented in Table 2.

DOI: 10.54254/2754-1169/59/20231010

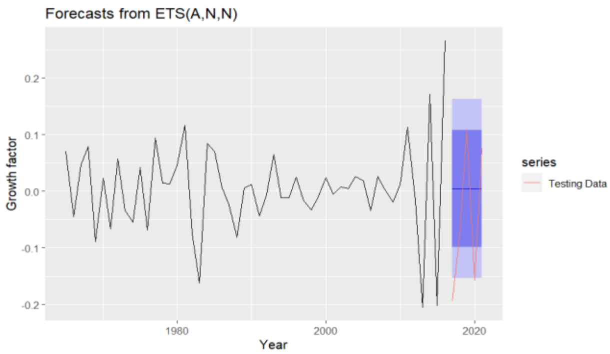

Figure 6: ETS (A, N, N) Model

Table 2: Summary for Accuracy Test of the ETS (A, N, N) Model

	ME	RMSE	MAE	MPE	MAPE	MASE
Training Set	-2.318538e-06	0.07921507	0.05447984	93.82874	97.06349	0.6207701
Test Set	-5.408744e-02	0.13370538	0.12458407	100.12727	100.12727	1.4195721

5.4. Five-year Forecast using ARIMA(0,0,1) Model

A comparative analysis of both models was based on the Root Mean Square Error (RMSE). The RMSE is given by:

$$e_t = y_t - \hat{y_t} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \tag{2}$$

For the ARIMA model, the RMSE was 0.063 for the training set and roughly 0.105 for the testing set. In comparison, the ETS model yielded RMSE values of 0.08 and 0.134 for the training and testing datasets, respectively (Table 3).

Given the comparative accuracy of the ARIMA (0,0,1) model, it was selected to generate birth rate projections for the next five years, with findings presented in Figure 7.

Table 3: Summary for Accuracy Test of the ARIMA (0,0,1) Model

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2022	0.00000000	-0.1079529	0.10795287	-0.1650997	0.165099668
2023	0.00000000	-0.1079529	0.10795287	-0.1650997	0.165099668
2024	0.00000000	-0.1079529	0.10795287	-0.1650997	0.165099668
2025	0.00000000	-0.1079529	0.10795287	-0.1650997	0.165099668
2026	0.00000000	-0.1079529	0.10795287	-0.1650997	0.165099668

DOI: 10.54254/2754-1169/59/20231010

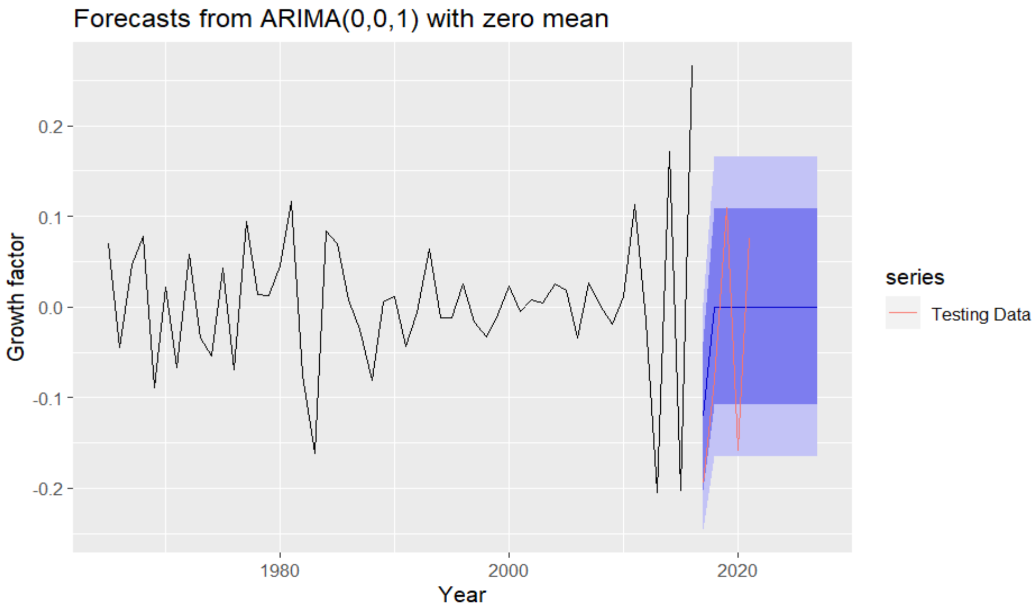

Figure 7: 5-Year Forecast using ARIMA (0,0,1) Model

6. Results and Analysis

6.1. Forecast Results Using ARIMA(0,0,1) Model

Our projections, derived from the ARIMA (0,0,1) model, as represented in Figure 6, demonstrate a uniformity in values for the point forecast column spanning the years 2022-2026. Notably, each year within this timeframe exhibits a value of 0.

6.2. Implications for Growth Factor and Growth Rate

Given the constancy in the forecasted differences year on year, the growth factor is anticipated to remain stagnant over the period of 2022-2026. The model predicts a growth factor consistent with 2021's recorded value of 0.8826291.

To discern the implications for the growth rate of the birth rate, one must consider the difference between the growth factor and unity. Specifically:

$$\text{Growth Rate of Birth rate} = \text{Growth Factor} - 1 \tag{3}$$

Employing the predicted growth factor, we deduce:

$$\text{Growth Rate of Birth rate} = 0.8826291 - 1 = -0.117 \tag{4}$$

Thus, the projections indicate a sustained decline in China's birth rate at a rate of 0.117 annually for the period 2022-2026.

6.3. Projections for Fertility Rate

Our analysis reveals a concerning trend for China. The persistent and marked decline in the birth rate suggests a challenging demographic scenario ahead. With an impending decline in fertility rates over the ensuing half-decade, China is poised to grapple with significant socio-economic ramifications, potentially extending well beyond the forecasted period [7].

7. Discussion

7.1. Data Transformations and Model Selection

In order to address the non-stationary character of the initial birth rate data, our methodology necessitated specific changes to ensure the stationarity of the time series. The utilization and juxtaposition of the ARIMA and ETS forecasting models were crucial in the process of selecting the ARIMA model as the preferred method for projecting China's birth rates for the upcoming five-year period commencing in 2021 [8].

7.2. Limitations

01. Exclusion of Data: Our decision to exclude birth rate data from 1960-1963, coinciding with the Great Leap Forward, may result in missing data points. This omission may have led to skewed calculations and conclusions due to the lack of a comprehensive dataset.

02. Metrics for Model Evaluation Our comparative evaluation of the ARIMA and ETS models during the training and testing phases was predominately based on RMSE values. While the root-mean-square error (RMSE) is a useful metric, additional evaluation parameters such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) could provide a more thorough model evaluation [9].

03. Analysis of Residuals: Our ARIMA model residuals exhibited a substantially high p-value. This raises questions about the consistency of these residuals with white noise, indicating possible model deficiencies.

04. Model Variability: Due to our academic familiarity with the ARIMA and ETS models, the scope of our study was limited to these methodologies. Despite the fact that the ARIMA model performed better than the ETS model in our tests, it may not be the ideal model for predicting future birth rates. Our limited exploration of alternative forecasting models is a result of our limited knowledge and experience [10].

8. Conclusion and Further Research Directions

8.1. Conclusion

Using a methodological approach to data transformation, our study demonstrates the validity of the log-transformed growth factor of the original birth rate data following a single differentiation. The statistical significance, as indicated by a p-value that is significantly less than the 5% significance threshold, demonstrates that our time series are stationary.

Our comparison of the ARIMA and ETS models on the training set demonstrates that the ARIMA(0,0,1) model is preferable for predicting China's birth rates. Curiously, this model's projections indicate a stagnant growth rate for the ensuing five years beginning in 2022. In practical terms, this implies that China's birth rate will continue to decline at an annual rate of 0.117. These findings shed light on the imminent crisis posed by China's falling fertility rates.

8.2. Future Research Directions

The gravity of the situation nudges us towards a multifaceted research agenda:

Examination of Influential Factors: It is crucial to comprehend the myriad of socioeconomic factors influencing birth rates. A thorough review of relevant academic literature from reputable academic archives would cast light on these determining factors.

Quantitative Analysis: Using datasets germane to the identified factors, a comprehensive examination employing time series models in R studio can elucidate the complex relationship

DOI: 10.54254/2754-1169/59/20231010

between these factors and birth rates. A study of this nature would not only assess the impact of these factors, but also determine birth rate variations resulting from changes in these factors.

Policy Recommendations: Faced with this declining fertility trend, it is imperative to devise robust policy interventions. Exploring academic discourse on the effectiveness of government initiatives such as maternity benefits and welfare incentives will be helpful. These insights would facilitate the development of strategies to combat and potentially reverse China's declining birth rate.

By engaging with this future research paradigm, we aspire to foster a comprehensive understanding of the low fertility conundrum and champion impactful measures to rectify it.

References

[1] *Kohler, H.-P. (2001). Social Interactions and Fluctuations in Birth Rates. Oxford University Press EBooks, 145–182.*

[2] *CAO, S., & WANG, X. (2009). DEALING WITH CHINA'S FUTURE POPULATION DECLINE: A PROPOSAL FOR REPLACING LOW BIRTH RATES WITH SUSTAINABLE RATES. Journal of Biosocial Science, 41(5), 693-696.*

[3] *Hasan, N. I. A., Aziz, A. A., Ganggayah, M. D., Jamal, N. F., & Ghafar, N. M. A. (2022). Projection of Infant Mortality Rate in Malaysia using R. Jurnal Sains Kesihatan Malaysia, 20(2), 23-36.*

[4] *Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.*

[5] *Wei, Y., Wang, Z., Wang, H., Li, Y., & Jiang, Z. (2019). Predicting population age structures of China, India, and Vietnam by 2030 based on compositional data. PLoS One, 14(4), e0212772.*

[6] *Karunanidhi, D., & Sasikala, S. (2023). Robustness of Predictive Performance of Arima Models Using Birth Rate of Tamilnadu.*

[7] *Yang, S., Jiang, Q., & Jesús S.J. (2022). China's fertility change: an analysis with multiple measures.*

[8] *Jain, G., & Mallick, B. (2017). A study of time series models ARIMA and ETS. Available at SSRN 2898968.*

[9] *Liu, Q., Charleston, M. A., Richards, S. A., & Holland, B. R. (2022). Performance of AIC and BIC in Selecting Partition Models and Mixture Models. Systematic Biology.*

[10] *Takayama, N., & Werding, M. (Eds.). (2010). Fertility and Public Policy. The MIT Press.*