Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/58/20230915

Model Analysis on the Birth Rate in China

Yongjie Chen[1,a,*]

[1]*Department of Statistics, University of California, Davis*
One Shields Avenue, Davis, CA 95616
a. yjcchen@ucdavis.edu
**corresponding author*

Abstract: This paper delves into the modeling and prediction of the birth rate in China, exploring the connection between China's economy and the birth rate using R. The study aims to find the best model to make a short-term forecast for the future birth rate in China. The research analyzes the birth rate data from 1950 to 2018 and identifies the factors that contribute to the decline in the birth rate in China. The study uses the ARIMA and ETS models to forecast China's birth rate and compare their performance. The research findings suggest that the ARIMA model outperforms the ETS model in forecasting the birth rate in China. The study culminates in emphasizing the importance of grasping the birth rate trends, particularly in crafting social policies, strategizing infrastructure development, and fostering economic growth. Policymakers should use the insights from this research to address the challenges of an aging population with a slowing economy in China.

Keywords: time series analysis, stationary, ARIMA model, ETS model, regression

1. Introduction

With a continuous decline in China's birth rate, consumption, labor markets, innovation, and other economic aspects have been influenced. As aging accelerated, there were more retired beings with fewer newborns, which increased social welfare pressures and costs and led to a tight labor supply. Meanwhile, China's economic growth has slowed. Thus, based on such a background, this paper researches the future trend of China's birth rate and its correlation with economic development.

George Brooke Roberts (1960) proposed a connection between the birth rate and the economy when he suggested that economic theories could be applied to the fertility rate analysis [1]. Butz and Ward (1979) researched the unusual countercyclical relationship between the birth rate and the economy in the U.S. during the 1960s and 1970s and gave several potential explanations with historical context for this special trend. John Cleland and Christopher Wilson (1987) criticized the traditional theories regarding fertility transition and argued that, in addition to socioeconomic development and urbanization, other factors like family planning and reproductive health intervention are also critical to the fertility transition. Allen C. Kelley and Robert M. Schmidt (1995) researched the multidimensional correlation between various demographic changes (fertility rate, mortality rate, and age structure) and economic development. Oded Galor and David N. Weil investigated the relationship between economic growth and fertility with great consideration of the gender income gap and gender equality.

This study delves into the intricate dynamics of China's birth rate, correlating it with the country's economic indicators. To achieve this, birth rate data was collected from 1950 to 2018 and meticulously analyzed, highlighting key factors responsible for its decline. Furthermore, the predictive capacities of both the ARIMA and ETS models are assessed to determine the more effective tool for short-term China's birth rate forecasting. The prime objective of this research underscores the pivotal role of understanding these trends, which have profound implications for shaping socioeconomic policies and infrastructure planning. The remainder of this paper is segmented into sections on data description, methodology, and results analysis, and the last section is the conclusion for the whole research.

2. Data

The primary data on China's birth rate was obtained from National Bureau of Statistics of China (http://www.stats.gov.cn/english/). The first dataset of China's birth rate (per 1,000 people) for time series modeling and prediction was made annually within a time range from 1963 to 2021 since this paper removed the great negative impact of the Great Leap Forward (from 1958 to 1962) and the Three Years of Great Chinese Famine (from 1959 to 1961) in our research.

Another integrated and comprehensive dataset containing China's birth rate, GDP, investment, consumption, exports, and imports was created through R where the raw data was also downloaded from the same website. Given this comprehensive dataset, we can investigate further the relationship between China's economy and the birth rate through regression analysis.

3. Methodology

ARIMA models blend AR and MA components with differencing, also a subset of linear regression models that attempt to use historical observations to forecast the future values of our response [1].

3.1. ARIMA Model

"ARIMA" is short for AutoRegressive Integrated Moving average, which is a combination of AR model, integration, and MA models in time series analysis [1].

An autoregressive model of order p, AR(p), utilizes previous observations of the dependent variable y, up to the p-th time step, as predictor variables. Here, p signifies the quantity of lagged data points in the model, ϵ is white noise at time t, c represents a constant term, and φs denotes the model parameters [2].

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \epsilon_t$$

Where x_t is described as a function of p past values, p represents the order of the autoregressive model, $\varphi_1, \dots, \varphi_p$ are themodel's parameters and ϵ_t signifies white noise.

Integration, also known as differencing, is to take d times to get a stationary series from the original non-stationary one. A time series is considered stationary when its statistical properties remain throughout the course of time.

For the first difference$\nabla = (1 - B)$:

$\nabla x_t = (1 - B)x_t = x_t - x_{t-1}$is a stationary series after being integrated into the first order.

Differences of order d are defined as:

$$\nabla^d x_t = (1 - B)^d x_t$$

Where x_t is a series, d is order of differences, and B is a backshift operator [2].

A moving average model employs an approach similar to regression analysis, which uses a linear combination of previous errors to predict future data points, with the following formula given below:

$$x_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

The mean of the series is represented by μ is, model's order is given by q, the model coefficients are indicated by $\theta_1, .., \theta_q$, and terms $\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-q}$ are the white noise error components.

Therefore, the final ARIMA(p,d,q) model is obtained by merging all three models mentioned above[1].

3.2. ETS Model

"ETS," short for Exponential Smoothing, is a time series forecasting model that utilizes the three components of Error (E), Trend (T), and Seasonality (S) to make predictions for future value based on past data. Each of these three can be typified as either "additive," "multiplicative," or "absent" [2]., The ETS (A, N, N) model was autonomously selected in this specific case, signifying uncomplicated exponential smoothing with additive discrepancies. The value of the smoothing parameter alpha stands at 1e-04, which is roughly equivalent to 0. According to Hyndman & Khandakar, when the value is 0, the series' magnitude remains static over time [3].

3.3. Root Mean Squared Error (RMSE)

Mean squared error is an estimate which measures the average squares of the errors

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Where n represents the total count of observations, Y_i stands for actual the response variable, and $\hat{Y}_i$ stands for the estimated response.[5].

Root mean squared error (MSE) is the square root of the value of MSE, with a formula given below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n}}$$

Where n represents the total count of observations, Y_i stands for actual the response variable, and $\hat{Y}_i$ stands for the estimated response.

3.4. Augmented Dickey-Fuller

the Augmented Dickey–Fuller (ADF) test is utilized to assess the assumption that a time series sample embodies a unit root. Depending on the specific variant of the test applied, the alternative hypothesis typically pertains to either stationarity or trend-stationarity. The ADF test is essentially an improved form of the Dickey–Fuller test, tailored for a broader and more intricate array of time series models [6].

DOI: 10.54254/2754-1169/58/20230915

4. Results

4.1. Time Series Analysis

Since the raw data on China's birth rate was non-stationary, we have made some transformations to the data to make the time series stationary. We first compute the growth factor and make it log-transformed to standardize the time series. However, we found that the log transformation failed to make further standardization to the original data of the growth factor. So, we made one more attempt at another differencing on the logarithm value of the growth factor, and, using the Augmented Dickey-Fuller Test, we successfully rendered this series stationary with a P-value smaller than 5%.

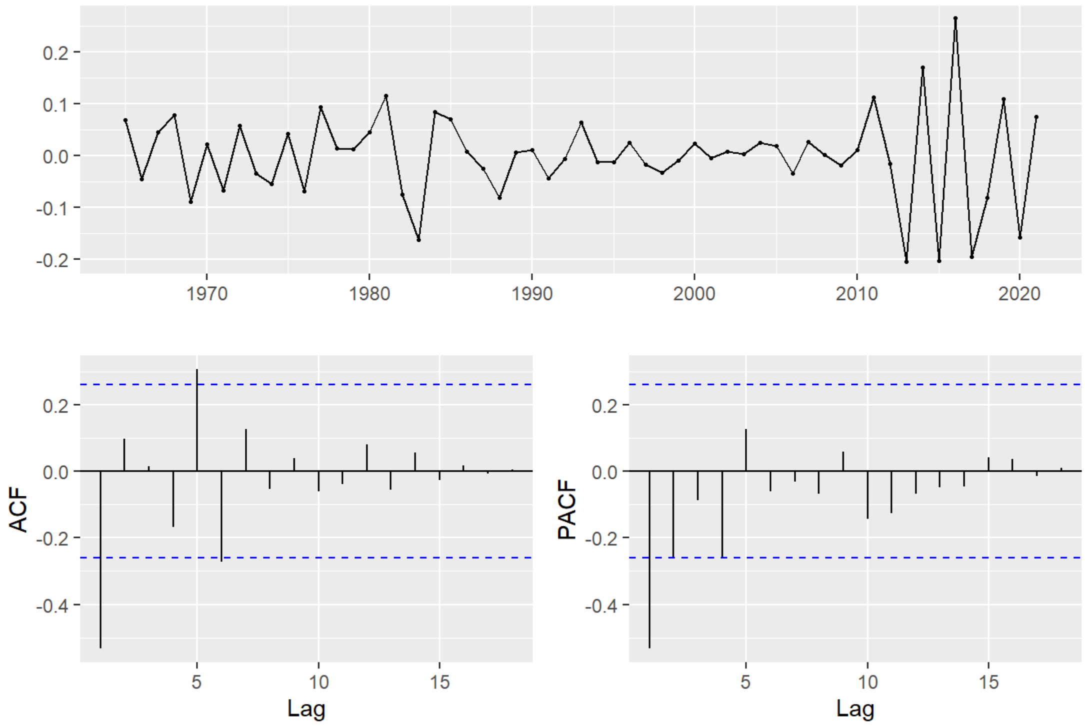

Figure 1: Time Series Plot After Transformed and Difference with its ACF and PACF.

Once transformed and differenced, the data was divided into two subsets: the training subset, used for data fitting and training, and the testing subset, utilized to assess the model's precision. Data spanning from 1964 to 2016 was assigned as the training set, while the data covering the period from 2017 to 2021 was selected as the testing set.

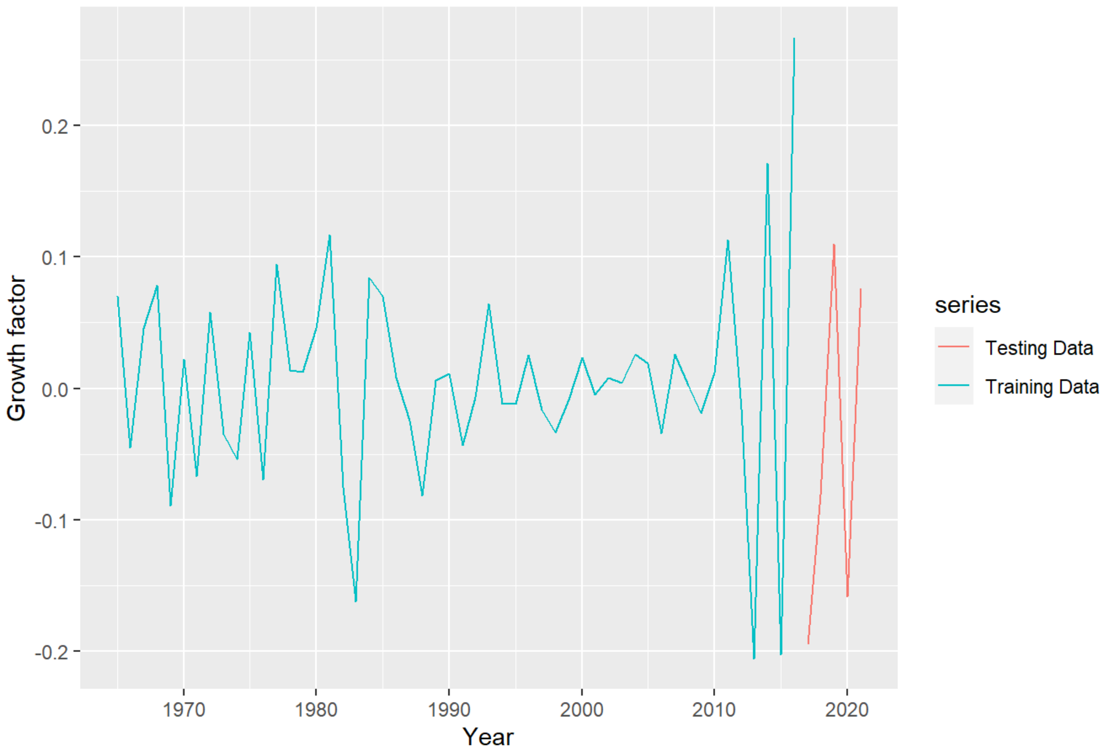

Figure 2: Time Series Plot Displayed by Training and Testing Data.

After getting the accuracy measured by RMSE, we can compare the ARIMA and ETS models to determine which model is more appropriate for predicting the future value.

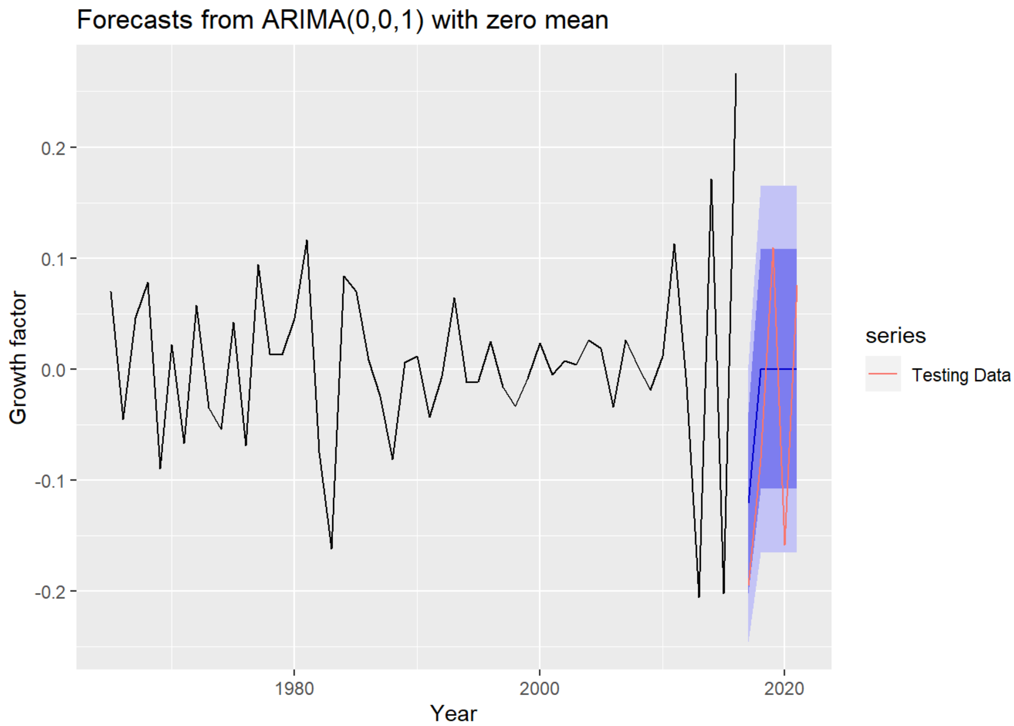

Figure 3: ARIMA (0,01) Model with A Five-Year Forcast.

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

	ME	RMSE	MAE	MPE	MAPE	MASE
Training	0.00959	0.0630	0.04647	97.99229	166.40961	0.52953
Testing	0.02558	0.10456	0.09956	87.60218	87.60218	1.134472

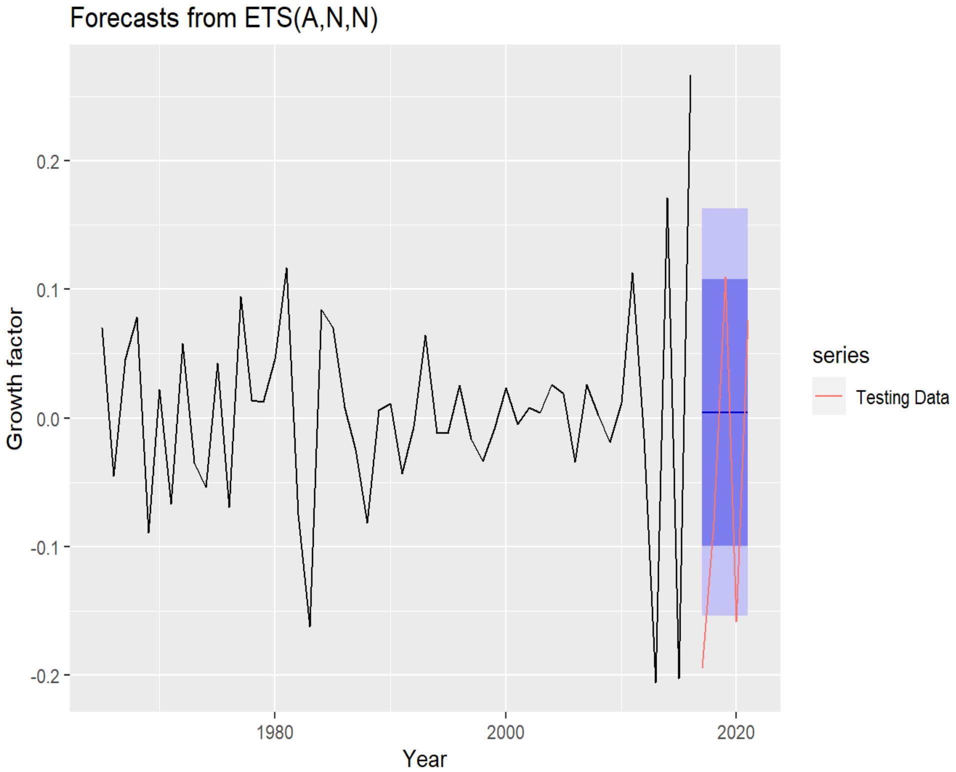

Figure 4: ETS(A, N, N) Model with A Five-Year Forcast.

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

	ME	RMSE	MAE	MPE	MAPE	MASE
Training	-0.00000231	0.07921507	0.054480	93.82874	97.06349	0.6207701
Testing	-0.054087	0.133705	0.124584	100.12727	100.12727	1.4195721

RMSE measures how well a model can predict data, with lower values of RMSE indicating better fit. It clearly reveals that the ARIMA model has a smaller RMSE value of 0.063 than ETS model does with RMSE value of 0.0792. In other words, ARIMA model in this case could provide a higher accuracy for the future value prediction on the China's birth rate.

4.2. Regression Analysis

Since we want to investigate further the link between China's birth rate and economic growth, another comprehensive dataset will be employed to perform the regression with the Ordinary Least Square Method. GDP was treated as a response and other variables of birth rate, consumption, investment, import, and export are factors. The summary of models is given below:

Table 3: Summary for Linear Models with Additional Variables.

	Model1	Model2	Model3	Model4	Model5
(Intercept)	1.49e-12*** (1293971)	0.0188* (-4.2e+04)	0.3875 (-1.728e+04)	1.87e-06*** (-8.886e+04)	3.51e-06*** (-8.801e+04)
Birth Rate	6.95e-10*** (-64619)	0.1307 (1.413e+03)	0.7524 (3.183e+02)	2.65e-05*** (3.756e+03)	4.06e-05*** (3.724e+03)
Consumption		>2e-16*** (3.813e+01)	3.893-14*** (3.201e+01)	<2e-16*** (3.324e-01)	<2e-16 (3.310e+01)
Investment			0.0316 (2.881e-01)	0.735 (-3.166e-02)	0.7709 (-2.775e-02)
Export				4.32e-09*** (8.891e-01)	0.7197 (-1.235e-01)
Import					0.0915 (1.119)
R-squared	0.6087	0.9981	0.9984	0.9993	0.9993

*‘***’ p<0.001, ‘**’ p<0,01, ‘*’ p<0,05, ‘.’ p<0.1, ‘’ p<1*

The estimate (coefficients) for each factor is in the parentheses in each cell.

It shows that the R-squared always increases as more variables are added each time the model is created. For the fifth model, which is the full model in this case, with all variables, only the birth rate factor is statistically significant with a positive coefficient. Since our research do stepwise modeling with additional one factors for each time, we can believe the summary of the last full model is valid and reliable. Thus, we can suggest that there is a highly significant correlation between the factor of China’s birth rate and its economy as measured by GDP value.

5. Conclusion

In conclusion, this research essay has demonstrated that the ARIMA and ETS models can be used to forecast the birth rate in China. The ARIMA model was found to be more accurate than the ETS model in terms of forecasting accuracy. The research findings suggest that the ARIMA model can be used to forecast the birth rate in China with high accuracy, which can help policymakers to address the challenges of an aging population in China. From the summary of all models fitted with stepwise regression, the last full model shows that only the birth rate emerged as statistically significant with a positive coefficient. This suggests a robust correlation between China's birth rate and its economic performance, as indicated by GDP value.

However, the research also has some limitations, such as the incomplete data in the dataset and the limited comparison of the ARIMA and ETS models with other models. Therefore, future research can focus on improving the dataset and comparing the ETS model with other models to determine the most effective model for forecasting the birth rate in China. Overall, this research essay provides valuable insights into utilizing time series to forecast China’s birth rate, which can inform policy decisions and contribute to the development of effective population policies.

References

[1] Roberts, G. B. (1960). Demographic and economic change in developed countries. NBER Books. https://ideas.repec.org/b/nbr/nberbk/univ60-2.html

[2] Butz, W., & Ward, M. P. (1977). The emergence of countercyclical U.S. fertility. The American Economic Review, 69(3), 318–328. https://eric.ed.gov/?id=EJ204725

[3] Cleland, J., & Wilson, C. C. (1987). Demand Theories of the Fertility Transition: An Iconoclastic View. Population Studies-a Journal of Demography, 41(1), 5–30. https://doi.org/10.1080/0032472031000142516

[4] *Kelley, A. C., & Schmidt, R. M. (1995). Aggregate population and Economic growth correlations: The role of the components of demographic change. Demography, 32(4), 543–555. https://doi.org/10.2307/2061674*
[5] *Galor, O., & Weil, D. (1993). The gender gap, fertility, and growth. https://doi.org/10.3386/w4550*
[6] *Newbold, P. (1983), ARIMA model building and the time series analysis approach to forecasting. J. Forecast., 2: 23-35. https://doi.org/10.1002/for.3980020104*
[7] *Jain, G., & Mallick, B. (2017). A study of time series models Arima and ETS. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2898968*
[8] *Abu Hasan, N. I., Abdul Aziz, 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 (Malaysian Journal of Health Sciences), 20(1), 23-36. https://doi.org/10.17576/JSKM-2022-2001-03*
[9] *Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice. OTexts.*
[10] *Mushtaq, R. (2011). Augmented dickey fuller test. SSRN Electronic.Journal.https://doi.org/10.2139/ssrn.1911068*