Prediction of China’s mortality rate under 5 in 2027 based on ARIMA model

Catherine Lin, Hannah Zhao, Sunny Sun, Tina Tian

G11 AP

Beijing National Day School

1. **Introduction**

Infant mortality rate (IMR) refers to the number of deaths of infants under one year of age per 1,000 live births in a given population. It is often used as an indicator of the overall health and well-being of a population. Society will further reduce child mortality through improvements such as health of expectant mothers and new moms. The research will focus on the prediction of China’s mortality rate under 5 in 2027 based on ARIMA model. And the result will 2. indicate the overall level of social development in China.

1. **Background**

IMR varies widely across different countries and regions, with some countries having very low rates while others have rates that are much higher. In general, factors that are associated with IMR include economic level, medical level, and educational level. Infant mortality rates have been declining globally in recent decades, due to improvements in maternal and child health care, as well as increased access to clean water and sanitation. However, there are still significant disparities in infant mortality rates between high-income and low-income countries, as well as within countries based on socioeconomic status and other factors.

There are three relevant pieces of information that help us better predict future mortality rates (under 5) based on related factors and test models. The economic level is the most important factor in the mortality rate. Higher values of GDP always results in lower mortality rate because GDP directly relates to the nation’s medical level and sanitation condition, which means people in the country have more resources and capital to take care of the kids, as well as the society can invest more in developing medical technologies. For infants, birth defects, preterm birth and low birth weight, sudden infant death syndrome, injuries, and maternal pregnancy complications are all main reasons for infant death, which contribute to children's mortality rate. As the medical level improves, sick children can receive a cure in time, and children with certain disabilities can be diagnosed earlier and treated under better care. Malnutrition reduces the effectiveness of health interventions, tainted water sources result in diarrheal illnesses, and unsanitary behaviors make mothers and children more susceptible to illness. The second resource used the autoregressive integrated moving average (ARIMA) model to forecast the infant mortality rate. They examined the accuracy of the model and concluded that there would be a decreasing trend of IMR (2017-2025). The third related resource specified is the ARIMA model. ARIMA model analyzes and forecasts non-stationary time series data, and it applies to abundant real-life situations when having a data set with non-linear patterns. These credible resources offer a comprehensive exploration of the accurate forecast for China’s mortality rate under 5 and implement the result to reduce the burden of the infant mortality rate.

1. **Method**

We first examine the stationarity of our data. To test whether or not it is stationary, we used ACF to observe the pattern of data.

Through the graph, we note that the mortality rate of child under 5 years old is non-stationary. So, after that, we tried to find its stationarity through differencing, which is by calculating the difference between successive observations. The result shows a first difference suitable for the mortality rate under 5.

Then we used the autoregressive model to estimate the value of p. The term autoregression indicates that it is the regression of a variable to itself. The formula of our AR(p) model, an autoregressive model of order p, is yt=c+ϕ1yt−1+ϕ2yt−2+⋯+ϕpyt−p+εt, where εt is the white noise.

The MA(q) Model, moving average model, is used to evaluate order q.

Through the process of finding p and q, it is sometimes hard to tell their values just simply by observing the time plot, so we can use ACF and PACF plots to do so. The ACF plot illustrates the autocorrelations, the relationship between two y values with a difference of k. And PCF measures the partial autocorrelations, the relationship between two y values after removing the effects of lags.

We can use AIC, AICc or BIC to test and help us to select the most suitable value of p and q.

1. **Result**

In this article, data of IMR from 1969-2021 in China is conducted based on real data, and the accuracy of model will be measured. The model shown in figure 1 is the IMR in China from 1969-2021.

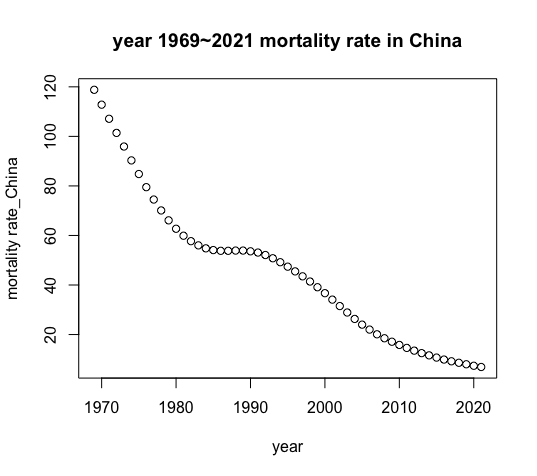
****

Figure 1: Infant Mortality Rate in China

Based on the Box–Jenkins approach, our study will be carried out in three parts: identification, estimation, and verification.

**Identification**

In this step, we are going to identify the conditions of the data for the model. We start with checking whether the data is stationary.

For stationary test, we use both Augmented Dickey-Fuller Test. The results is shown in Table1.

|  |  |  |  |
| --- | --- | --- | --- |
| Differential times | 0 | 1 | 2 |
| P-value | 0.0767 | <0.01 | 0.6625 |

Table 1: ADF Test results

When the time series is after one difference, the resulting p-value is smaller than 0.01, which is smaller than the significance level 0.05. As a result, we reject the null hypothesis and accept the alternative hypothesis, which shows that the data is stationary. Figure 2 shows the model after one difference.

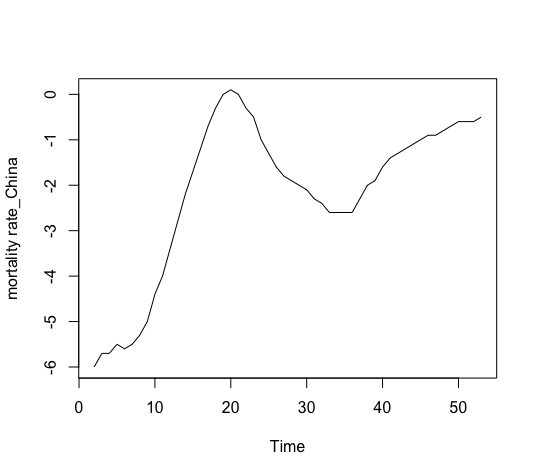


Figure 2: Data after one difference

After verification of the stationary of the series, we notice from the ACF and PACF correlation that our model is not pure AR or pure MA. We therefore tested several models to identify the most suitable one for sales. In figure 3 and 4, graph of ACF and PACF are shown.

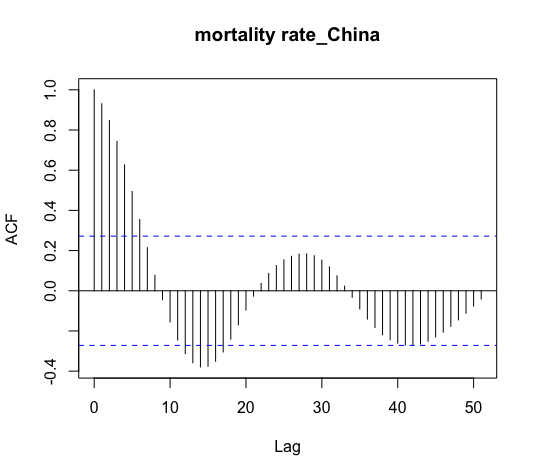


Figure 3:Aautocorrelation function graph

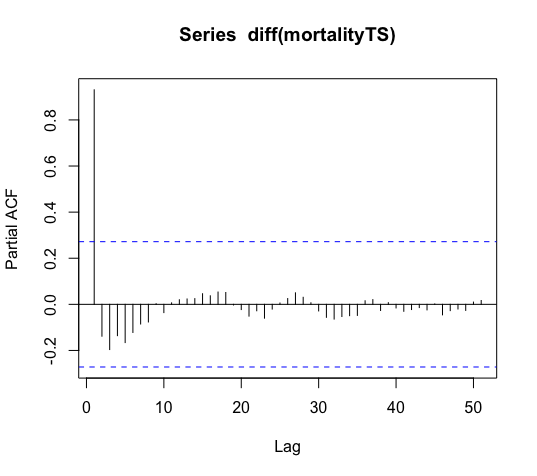


Figure 4: Partial autocorrelation function graph

Next, we need to figure the white noise of the sequence, and we use Ljung-Box test. The result of the test is in Table 2, which showed that the sequence is non-white noise and can be fitted.

|  |  |
| --- | --- |
| X-squared | 47.626 |
| df | 1 |
| P-value | 5.157e-12 |

Table 2: Ljung-Box test result

**Estimation**

Then, we use graph of ACF and PACF to determine the parameters p and q.

The ARIMA we used is (7,1,1). Table 3 summarizes the data of the model and the value of all the coefficient.

ar1 ar2 ar3 ar4 ar5 ar6 ar7 ma1

0.6963 1.4019 -0.5864 -0.8547 0.1081 0.1598 0.0703 0.9997

Table 3: Coefficient of the model

**Verification**

The accuracy of the model is estimated by Ljung-Box test and residual of the model.

P-value for Ljung-Box test to check if the sequence residuals are white noise is 0.8078, which we won’t reject the null hypothesis.

Figure 5 and 6 shows acf and pacf of the residuals.

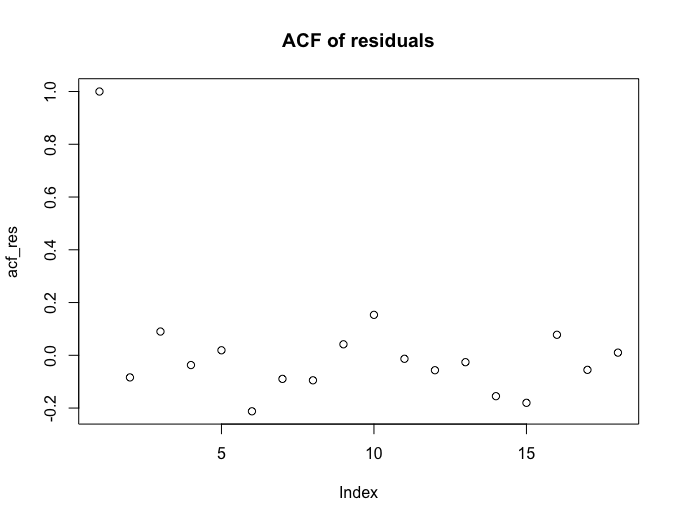


Figure 5: Aautocorrelation function graph for residuals

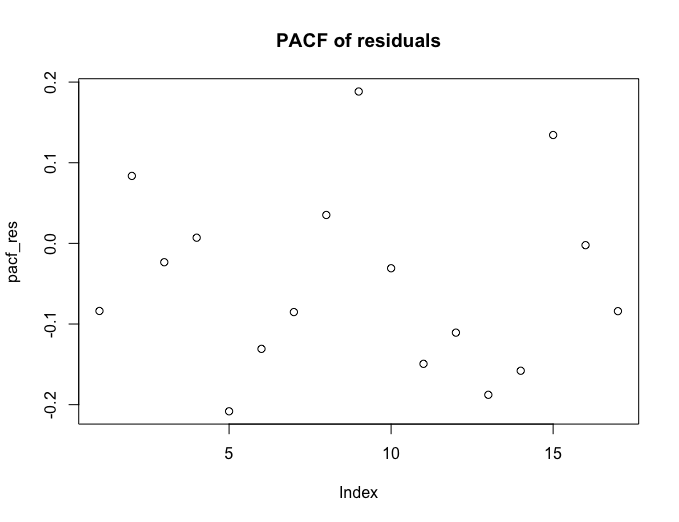


Figure 6: Partial aautocorrelation function graph for residuals

**Forecast**

Figure 7 and Table 4 shows the final forecast using the ARIMA model.

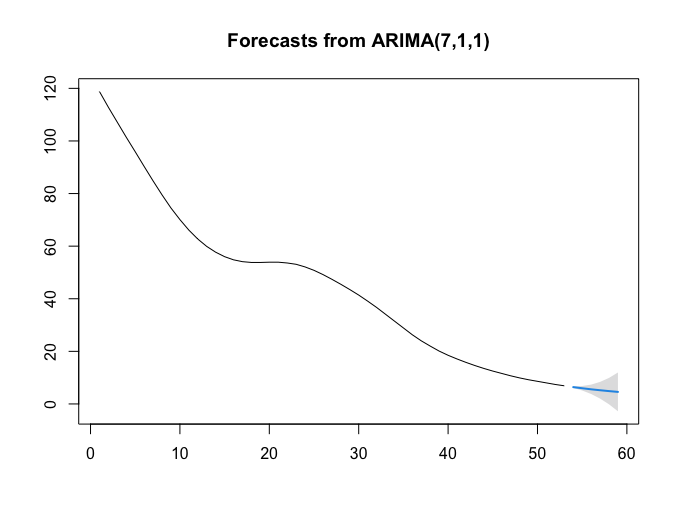
****

Figure 7: Forecasts of data

Point Forecast Lo 99 Hi 99

54 6.393939 6.0960473 6.691831

55 5.972291 5.1207025 6.823880

56 5.565297 3.7849554 7.345639

57 5.211960 2.0420834 8.381836

58 4.883018 -0.1317681 9.897804

59 4.580872 -2.7931401 11.954884

Table 4: Forecasts and confidence interval of 99%

Based on the selected ARIMA model, the IMR was forecasted for postsample period of 9 years (2017-2025). The results showed a declining trend of IMR in the predicted period (figure 7). The predicted IMR in China is 4.58 in 2027.

1. **Conclusion**

Technical analysis provides compelling evidence of a downward trend in China's IMF, thus supporting our hypothesis. It can be seen that ARIMA model has a good prediction effect on IMR.

1. **Discussion**

Limitations of the model lie in the strict requirements for large sample size and stationary time series, and the ARIMA model can only be used to predict linear relationships. Furthermore, while the accuracy of predictions has shown high efficacy, improving the accuracy of models to near-perfect predictions remain the main goal of countless studies over the past decades. Admittedly, the ARIMA model still needs to be further improved in the future. The data we have so far are only close to the truth, and give us only a modest idea of the future trajectory of child mortality rate in China.

In addition to the shortcomings of the ARIMA model itself, other factors may negatively affect the results as well. For example, the cohort effect may be at work in the process because of the long time series required. As a result, previous data may contribute little to help predict outcomes, potentially undermining the predictability of the ARIMA model or even shifting the analysis in the opposite direction.

1. **Reference**

[1] United Nations. (n.d.). Education Is Key to Reducing Child Mortality: The Link Between Maternal Health and Education | United Nations. https://www.un.org/en/chronicle/article/education-key-reducing-child-mortality-link-between-maternal-health-and-education

[2] Infant Mortality | Maternal and Infant Health | Reproductive Health | CDC. (n.d.). https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm

[3] Mishra, A., Sahanaa, C., & Manikandan, M. (2019). Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model. PubMed, 26(2), 123–126. https://doi.org/10.4103/jfcm.jfcm\_51\_18

[4] Brwnlee, J. (2020). How to Create an ARIMA Model for Time Series Forecasting in Python. MachineLearningMastery.com. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/#:~:text=A%20popular%20and%20widely%20used,structures%20in%20time%20series%20data.

[5] Mishra, A., Sahanaa, C., & Manikandan, M. (2019b). Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model. PubMed, 26(2), 123–126. https://doi.org/10.4103/jfcm.jfcm\_51\_18

[6] Adhikari, R., & Agrawal, R. K. (2013). An Introductory Study on Time Series Modeling and Forecasting. In arXiv (Cornell University). Cornell University. <https://arxiv.org/abs/1302.6613>

[7] CME Info - Child Mortality Estimates. (n.d.). https://childmortality.org/data