ISSN 2616-7433 Vol. 3, Issue 6: 68-71, DOI: 10.25236/FSST.2021.030611

Number regression and prediction of Chinese seniors based on the ARIMA model

Yingshuo Li[1], Yu Liu[2]

[1]*Qingdao University, Qingdao, Shandong, 266000, China*
[2]*Weihai Campus of Shandong University, Weihai, Shandong, 264200, China*

Abstract: *China's aging population developed rapidly in the 21st century, It has brought a significant impact on the changes in China's social security system and labor force structure, In 2021, the "seven general" data show that the aging process of China's society continues to climb, This paper will be based on the number of 65 in China from 1990-2020, Using the time-series ARIMA model, , Differential ideas derived from data stationarity, Unit root inspection, Model order, AIC and BIC coefficients in contrast, Model error analysis, As well as model prediction analysis, the number of elderly population in China from 2021 to 2030, The research results show that the number of elderly people in China is still increasing in the next few years, And put forward the corresponding policy suggestions for the results.*

Keywords: *Time series, ARIMA model, Population aging, Population prediction*

1. Introduction and Literature review

Population aging refers to the reduced population fertility rate and per capital life extension caused by the number of young population, the older population increases, resulting in the proportion of the elderly population, a country or region over 60 years old elderly accounts for 10% of the total population, or over 65 accounted for more than 7% of the total population, the country or region is in the aging society, Lu Jiehua, Liu Qin. [1] Based on the data of "seventh national census", we found that affected by various factors such as declining fertility level, the proportion of elderly population continued to rise, the total population of 65 and above was 191 million, accounting for 13.50% of the total population, up 4.63 percentage points compared with 2010, China's aging social form further deepened,

Domestic scholars also conduct in-depth research on the population aging process in China from the multi-dimensional research direction of population policy, data prediction and economic influence. Taurus [2] According to the University of Washington population forecast in 2020, for the first time outlined the basic trend of negative world population growth in the 21st century, compared with the United Nations population forecast in 2019, low fertility level is the main mechanism of negative population growth in the 21st century. Sun Ting and Wang Xueqin [3] Using clustering method and coword analysis of policy text to systematically and comprehensively comb my aging policy, and said that with the deepening of China's strategy of coping with aging countries, China will expand the aging policy research in policy coordination, policy implementation and policy evaluation in the future. Cao Yanan [4] Based on the Logistic population model, the number of elderly population aged 65 from 2020 to 2020 predicted the data aged 65 and above from 2021 to 2040. The results show that the number of aging population in China still showed a trend of continuous rise.Yan Yujun, Yan Yunlou [5] Based on GM(1,1) and principal component analysis, the degree of aging in Shanghai will intensify in the next 20 years, and the principal component analysis method found that the three factors affecting the aging in Shanghai are mainly social, policy and economy. Comprehensive consideration, the future aging population forecast, and adjust policy and economic structure will be a big problem in China and even the world, ARIMA model has simple model, and does not need to use the advantages of other exogenous variables, while few scholars use the ARIMA model of Chinese elderly population analysis and prediction, this paper will use ARIMA model in the model stability test to predict aging population in 2021-2030.

2. Data sources and the theoretical model

(1) Data source and preprocessing

ISSN 2616-7433 Vol. 3, Issue 6: 68-71, DOI: 10.25236/FSST.2021.030611

The data in this paper are from the National Statistical Yearbook. It selected the number of 65 seniors (unit: ten thousand people),from 1990 to 2020.The raw data was visualized and visually detected as shown in Figure 1, and the time series map showed an obvious non-stationary trend. In order to ensure the accurate data processing, for the unit root test of the raw data, the p value is greater than 0.05, the significance condition was not met, and the stationary null hypothesis was rejected to prove that the sequence is a non-stationary sequence,

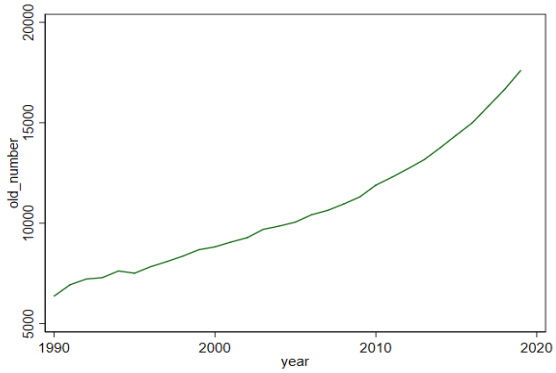

Figure 1: Number of elderly population in China from 1990-2020

(2) Model theory and fixed order

ARIMA model is called autoregressive integrated sliding mean model, abbreviated A RIMA (p, q, d), is one of the important models to fit and predict time series, in ARIMA (p, d, q), AR represents "autoregressive", p as autoregressive terms, MA represents "sliding mean", q for sliding mean terms, d to make difference number to make it stationary sequence, AMIRA (p, d, q) model:

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d X_t\left(1+\sum_{i=1}^{q}\theta_i L^i\right) \quad (1)$$

Where the L is the lag operator, and$d\epsilon Z$,

Figure 1 shows that the number of Chinese aging population continues to rise, which is a clear non-stationary time series, using

Differential idea and unit root test method, the first order difference to the data, if the data still does not meet the stability, the second order difference, and so on, repeated, until the data is processed as a stable time series data.The differential formula is as follows

$$\nabla x_{i=} x_i x_{i-1} \quad (2)$$

$$\nabla^p x_i = \nabla^{p-1} x_i - \nabla^{p-1} x_{i-1} \quad (3)$$

The first order difference equation and unit root test, the timing of the stationary time series, as shown in Figure 2, the difference in the second order, in the results of Z value is less than the critical value of each test, and p value is 0.0000, less than 0.05, do not reject the original hypothesis, meet the significance condition, so determine the parameter d value is 2

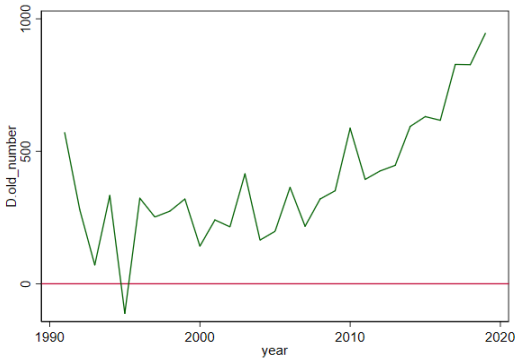

Figure 2: Order difference

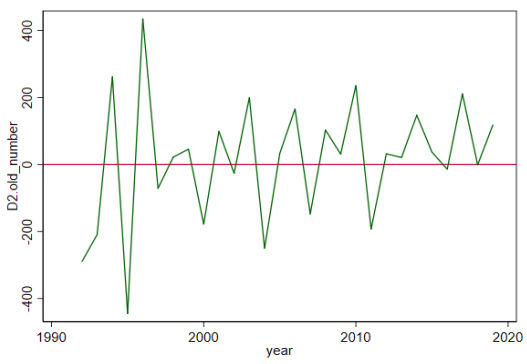

Figure 3: And second order difference

Plot the AC(autocorrelation) to determine the P values, PAC(partial self-phase)) Fig determines q, As can be seen in Figure 4 and Figure 5, ACF shows a distinct dragging tail after the second order, PACF showed significant truncation after the third order, The relatively appropriate sequence after the

ISSN 2616-7433 Vol. 3, Issue 6: 68-71, DOI: 10.25236/FSST.2021.030611

difference is ARIMA(2,2,3), And here at p, q, Using the AIC(red pool information criterion) and the BIC(Bayesian information criterion), the smaller the AIC and BIC, the better the criterion, To screen for the parameter values of the optimal ARIMA model, The resulting optimal ARIMA model is the ARIMA(2, 2,3) The calculation results are shown in Table 1 below:

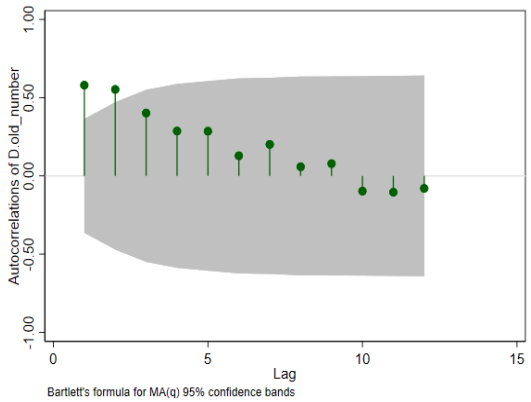

Figure 4: Sequence auto correlation

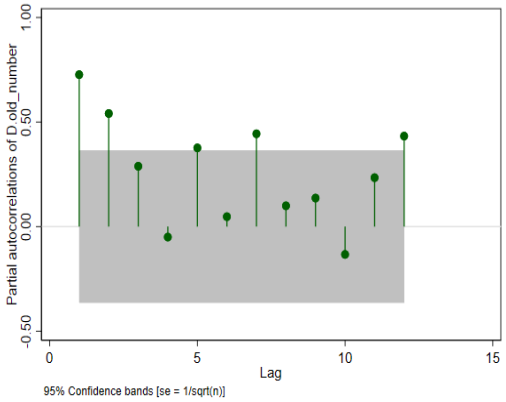

Figure 5: Sequence partial autocorrelation

Table 1: AIC and BIC tests

(p,q)	AIC test	BIC test
(2,3)	236.8481	241.5703
(1,3)	373.4463	381.4395
(2,2)	251.6749	255.65

(3) White noise test of the residual sequence

After modeling the time series, the residual value is predicted and white noise test, if the residual sequence is white noise sequence, indicating that the original sequence has full valuable information has been extracted, we can propose the following hypothesis:

$$H_0: p_1 = p_2 = \ldots = p_m = 0, \forall m \geq 1 \tag{4}$$

$$H_1: \text{At least one exists} p_k \neq 0,\ \forall m \geq 1, k \leq m \tag{5}$$

The white noise test for residual value can test the white noise using the Q statistic and LB statistic, n is the number of sequence observation periods, m is the specified delay period, j is the delay order, the closer the number Q statistic to 0, the greater the likelihood of white noise, and the Q statistic is as follows,

$$Q = n\sum_{j=1}^{m} p_j^2 \tag{6}$$

$$LB = n(n+2)\sum_{j=1}^{m} \frac{1}{n-j} p_j^2 \rightarrow X_{1-\alpha}^2(m) \tag{7}$$

Through the results of the white noise test, the test points all fall within a certain range, and the model has a good fitting effect.

(4) Model fitting effect test and prediction

Fit curves of the original and predicted sequence values were compared showing good fit as shown in Figure 6:

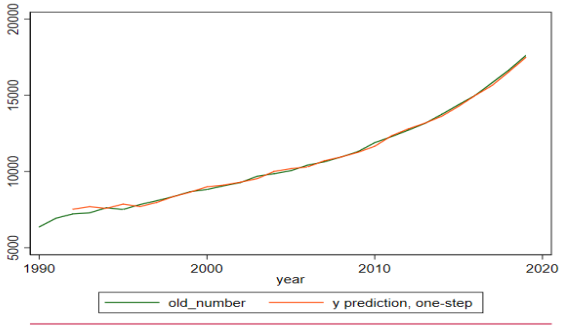

Figure 6: Comparplots of raw and prediction data

ISSN 2616-7433 Vol. 3, Issue 6: 68-71, DOI: 10.25236/FSST.2021.030611

Using the obtained ARIMA(2,2,3) model to predict China's aging population in 2021-2030, through prediction, you can see that the number of the elderly population still shows a growing trend year by year, social aging pressure, and late marriage and late childbearing social phenomenon also leads to the aging process in China, which also fits ourl conditions (unit: ten thousand people), predited value in table 8below:

Table 2: Population number is predicted for the next decade

Year	Elderly population number forecast value (ten thousand people)
2021	19708
2022	21060
2023	22413
2024	23985
2025	25578
2026	27423
2027	29320
2028	31424
2029	33674
2030	36051

According to the data of the seventh census in 2021, China's population aged 65 and above is 190.64 million, which is not much different from the forecast data. Through the above data, the aging population in China will account for 360.51 million in 2030, about to reach a quarter of the total population. It is imperative to adjust the population structure.

3. Conclusion and policy suggestions

In this paper, the number of elderly population is predicted by ARIMA model, And through the data stationarity, Unit root inspection, Model order, Model error analysis, The model prediction and analysis of the elderly population in China from 2021-2030, It is found that the number of elderly people in China is still increasing in the future, If the policy adjustment is not made in time, This would increase the burden on the existing working population, The increase in total spending on basic endowment insurance will also increase the government's financial burden.Based on the above conclusion, the following suggestions are put forward for improving the aging situation in China:

1) Alleviates the continuous consequences of policy constraints, fully opens and encourages childbirth, reduces the promotion of suitable women, work pressure, creates a perfect prenatal and postpartum security service system, provides appropriate government subsidies, and reduces the cost of childbirth and raising.

2) Promotes the development of the elderly industry, improves the elderly security system, reduces the social phenomena such as high-end elderly products and high-price nursing homes, and focuses on the old-age services and construction of the elderly groups in low-income elderly groups, especially the vast rural elderly groups, so as to realize the real "care for the elderly".

3) Improves the retirement system, implements the flexible retirement system, implements the differentiated retirement system for different positions and types of labor groups, and encourages enterprises and public institutions to implement re-employment policies for professional and technical personnel and field experts.Reduce the waste of talent phenomenon.

References

[1] Lu Jiehua, Liu Qin.Characteristics, influence and coping strategies of new forms of Chinese aging society —— is based on the interpretation of "seven pu" data [J]. Population and Economy, 2021 (05): 13-24.
[2] Original New, Taurus.Outlook and impact response of negative world population growth [J].Proceedings of Hebei University (Philosophy and Social Sciences Edition), 2021, 46 (01): 82-91.
[3] Sun Ting, Wang Xueqin.—— Policy Text Analysis Based on 1949-2020 [J].Chongqing Social Sciences, 2021 (09): 38-52...
[4] Cao Yanan, Cui Yujie.Study on the prediction of Chinese population based on Logistic population model [J]. Modern marketing (Scholastic Edition), 2021 (08): 192-193.
[5] Yan Yujun, Yan Yun Lou.Aging trend and influencing factors —— based on GM(1,1) and principal component analysis [J].Chinese Journal of Gerontology, 2021,41 (14): 3093-3098.