

# Prediction of the Situation of Postgraduate Entrance in China: Applicants and Enrollments

Chiyang Zhong \*

University College, University of Oxford, Oxford, The United Kingdom

\* Corresponding author: chiyang.zhong@univ.ox.ac.uk

**Abstract.** Applying for a postgraduate program has always been a popular choice after the four undergraduate years in China because it is deemed to increase competitiveness in the job market. In recent years, applying for a postgraduate program has become more and more of a necessity when entering the job market. Therefore, it is important to understand and predict the postgraduate entrance trend. This paper used ARIMA model to predict the number of postgraduate applicants and enrollments in the near future after using KPSS tests to determine the order of differencing, and a combination of unit root tests, and optimization of AICc and MLE to find the optimal order of autoregression and moving average. The result reached is that the number of postgraduate applicants follows an ARIMA (1,2,0) model, and that of enrollments follows an ARIMA (3,2,0) model. The number of both postgraduate applicants and enrollment will keep on rising in the near future, while the admission rate is mostly stable.

**Keywords:** ARIMA, postgraduate, applicant, enrollment.

## 1. Introduction

Applying for a postgraduate program has long been a tempting option of further study. From undergraduates to employees, many people have considered postgraduate program for a variety of reasons, no matter it is to further their interest, enhance their academic abilities, expand their social circle, or to increase their competitiveness when applying for jobs [1]. This is especially the case in China: the number of applicants for Postgraduate Entrance Exam has almost doubled from 2017 to 2021, and has increased over 10-fold since the start of the century [2]. Predictions also showed that this steady growing trend will be continued in the near future [3, 4]. Therefore, it is quite worthwhile to take a deep dive into the trend of the situation of postgraduate entrances in China, including the number of applicants and final offers, to find out how it might evolve.

Studying the trend of postgraduate entrances is not only motivating but also meaningful. In the most straightforward aspect, the study could provide a reference to prospective postgraduate applicants in China. The study could help them get a more statistically complete view of the situation of postgraduate application, and it would contribute as a factor for consideration when deciding whether to apply or not. The study would also be helpful to employers and companies in China: from the study they could have a better understanding of the how much of the students are willing to further their academic life, and how much of them are actually getting an offer. In the future they would be able to have a more down-to-earth expectation of the educational qualification and degree of those applying for their organization or firm and, based on the analysis and predictions, make or adjust their hiring strategies. There is also takeaway for higher institute in China. Given the increasingly tight educational resources in the higher institutes in China [5, 6], the analysis of the domestic trend of postgraduate applicants could provide a macroscopic view of the applying situation, and, by combining the past trend of the postgraduate applicants, the institute can decide on whether they should change their policies on admission or consider gaining or reallocating educational resources.

This paper aims to predict the future trend in the number of postgraduate applicants and student enrollment in the near future. In the analysis, the paper will use the Autoregressive Integrated Moving Average (ARIMA) model to predict both the scale of postgraduate applicants and admissions in the future. It is worth pointing out that there are existing papers and articles on this topic [7, 8], and different methods of forecasting are used, including QPSO-BP, Gray Forecast Model, and also the

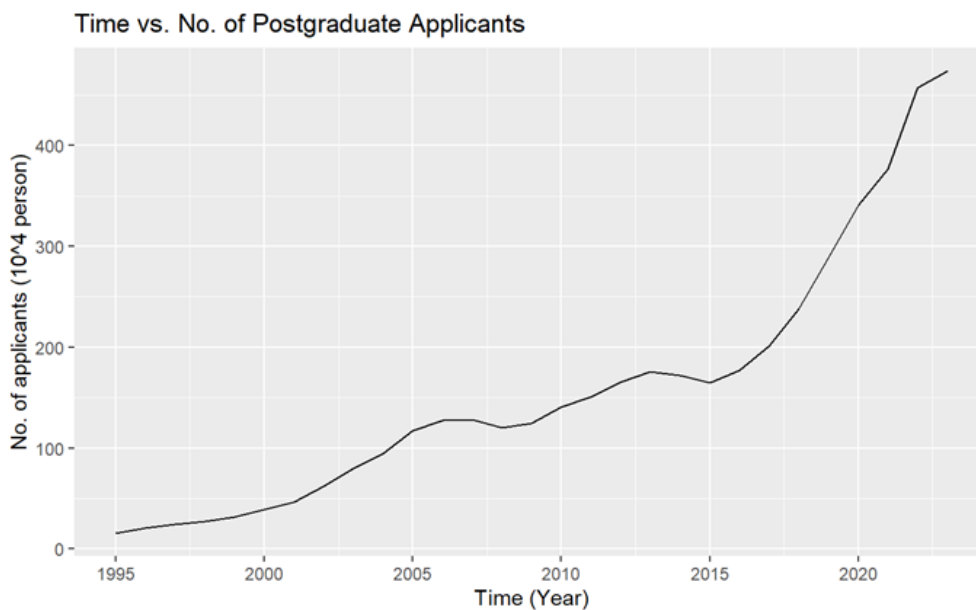
ARIMA model this paper uses. These models have their benefits, for example Gray Forecast Model is very useful when it comes to small-size data. However, there are also downsides: it is also the case that Gray Forecast Model is relatively accurate when dealing with exponentially evolving data and forecasting in short term, which are pretty strong assumptions. With influential policies in place in 2017, and the impactful COVID-19, it is possible that the predictions in these papers would no longer be valid, especially with the ARIMA model case, since there could be change point occurring due to fluctuation of GDP, fierce competition tied with employment, change of policies, and social situations [9, 10]. All these reasons only make it more necessary to study the scale and situation of postgraduate entrances in China.

## 2. Methods

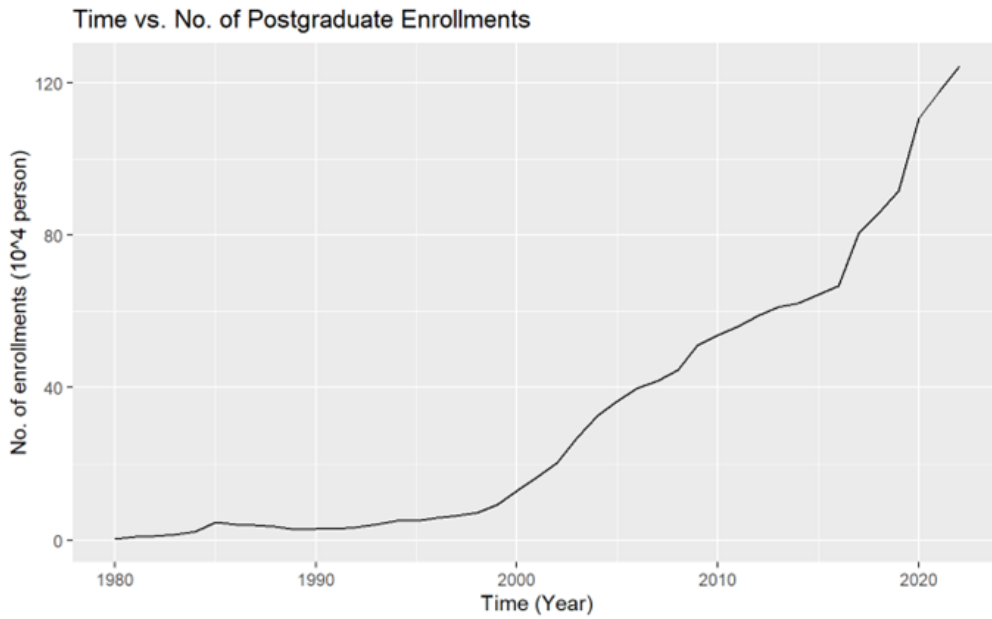
### 2.1. Data Source

The data of both the number of postgraduate applicants and that of enrollments are mostly collected from the Dydata website, and is in turn collected from eol.cn and the Ministry of Education of the People’s Republic of China. The data of number of applicants ranges from 1995 to 2023 with a total of 29 observations. The data of number of enrollments ranges from 1980 to 2022 with a total of 43 observations.

Fig 1 illustrates the number of postgraduate applicants in China each year from 1995 to 2023, and Fig 2 illustrates the number of postgraduate enrollments in China each year from 1980 to 2022. Qualitatively, the data shows that there is an accelerating increase trend in both the number of postgraduate applicants and postgraduate enrollments in China, and there does not seem to be periodicity.



**Figure 1.** Time (Year) vs. number. of Postgraduate Applicants in China from 1995 to 2023 (unit: 10<sup>4</sup>)

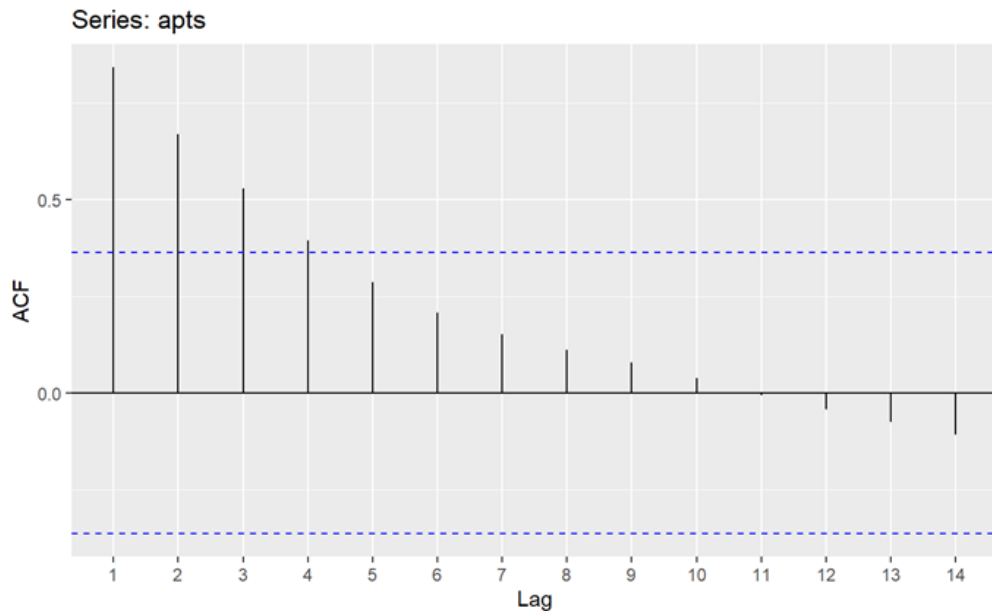


**Figure 2.** Time (Year) vs. number. of Postgraduate Enrollments in China from 1980 to 2023 (unit: 10<sup>4</sup>)

## 2.2. Data Processing

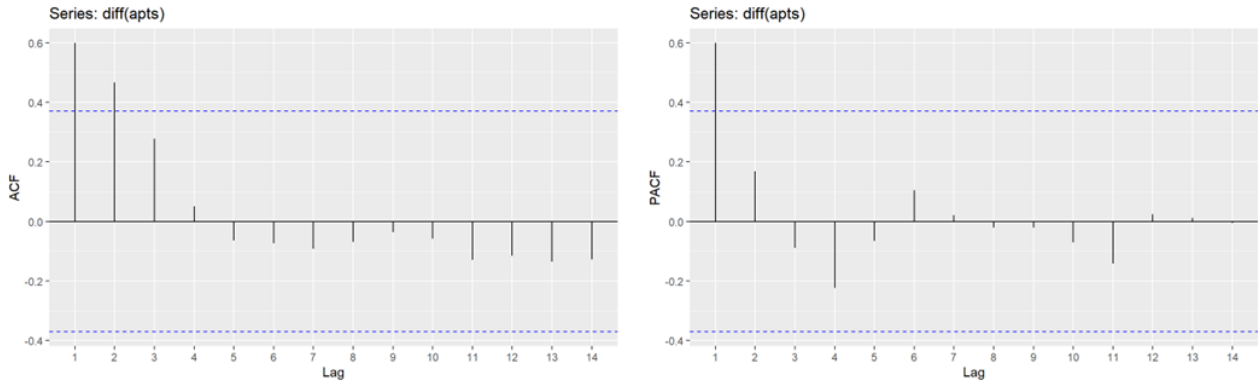
### 2.2.1. Applicant data

The following figure, Fig 3, is the ACF plot for the data. The figure shows that there is strong autocorrelation in the data. We can difference the data to try and counter it.



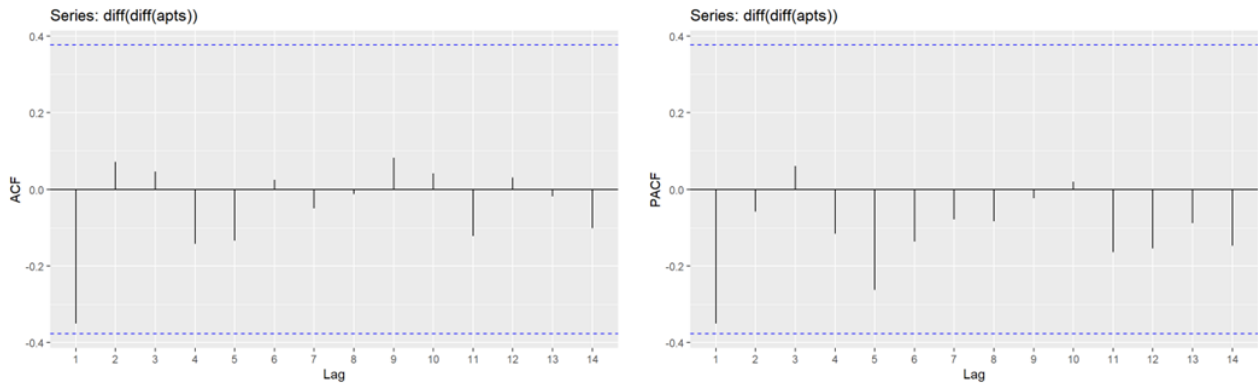
**Figure 3.** ACF plot of postgraduate applicants in China

The left figure of Fig 4 is the ACF plot of the first-order differenced data. In the figure, it can be observed that the first two strip exceeded the critical values. This pattern can also be demonstrated this using the PACF plot, which is presented in the right figure of Fig 4.



**Figure 4.** ACF (left) and PACF (right) plot of first-order differenced postgraduate applicants

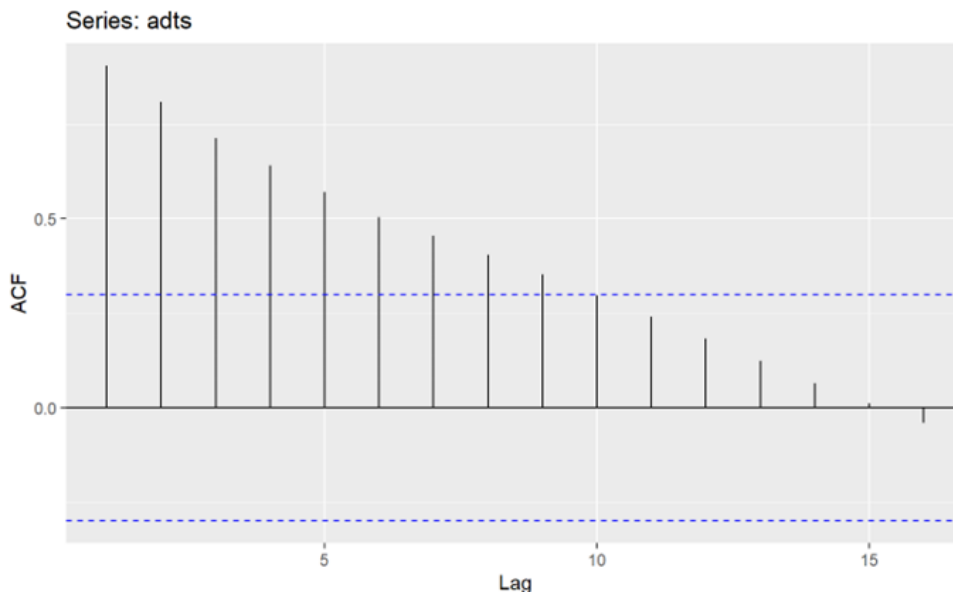
By a series of Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, we find that we would need to difference the data a second time to obtain stationarity. For the second difference, we find that both ACF and PACF are bounded in the critical range. This is demonstrated in Fig 5.



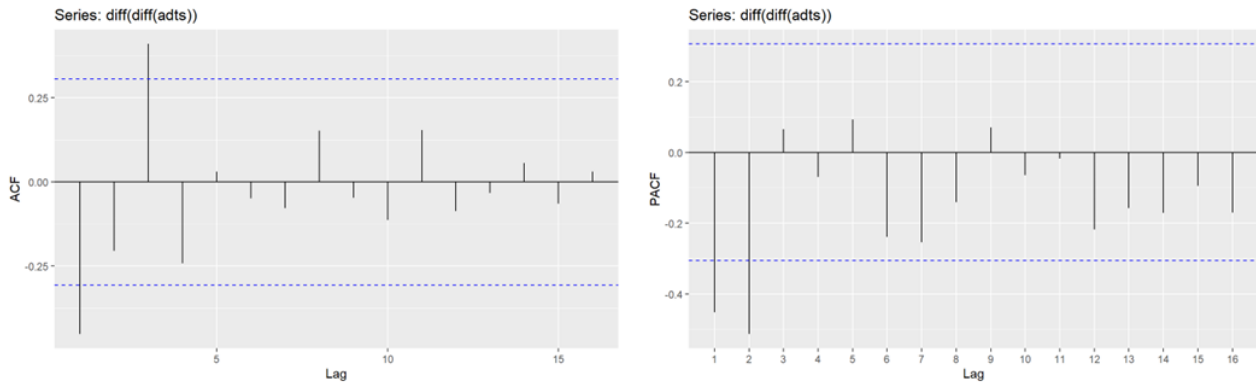
**Figure 5.** ACF (left) and PACF (right) plot of second-order differenced postgraduate applicants

### 2.2.2. Enrollment data

We follow a similar procedure for enrollment data. By a series of KPSS tests, we find that we would need to difference the enrollment data twice to obtain stationarity. We hence only illustrate the ACF and PACF plots for the non-differenced and first-differenced data in Fig 6 and 7. For the second difference, we find that ACF as first and third strip exceeding the critical range, and PACF has first and second strip exceeding the critical range. This is demonstrated in Fig 7.



**Figure 6.** ACF plot of postgraduate enrollments



**Figure 7.** ACF (left) and PACF (right) plot of second-order differenced postgraduate enrollments

### 2.3. Model Description

This paper will use the Autoregressive Integrated Moving Average (ARIMA) model for analysis of both the applicant and enrollment data. ARIMA model is a powerful and commonly-used tools when analyzing time series. One of its aims is to describe autocorrelation of the data. ARIMA models consist of three key components: autoregression (AR), integration (I), and moving average (MA). It is reflected in the ARIMA models as parameters: an ARIMA (p, d, q) model has p, d, q as parameters for the three components. In both applicant and enrollment data, we differenced the data twice, hence d=2 in both cases, as we differenced twice using KPSS tests.

### 2.4. Model Selection

By a combination of unit root tests, optimization of the corrected Akaike’s Information Criterion (AICc) and maximum likelihood (MLE), the model ARIMA (1,2,0) is obtained as the best model for the applicants’ data (Table 1), and that for the enrollments data is ARIMA (3, 2, 0).

We can compare some variations of the models and their corresponding AICc and root mean square error (RMSE) to verify that these are indeed good models (Table 2): The model for applicants’ data has the lowest AICc among the selected models, and its RMSE is only higher than ARIMA (2,2,1). The model for has the lowest AICc among the selected models, and its RMSE is only higher than ARIMA (3, 2, 1).

**Table 1.** Different ARIMA models and their corresponding AICc and RMSE

Model	AICc	RMSE
ARIMA (1,2,0)	228.49	14.64666
ARIMA (0,2,0)	232.82	16.72254
ARIMA (1,1,0)	240.36	15.81309
ARIMA (2,1,0)	239.5	14.72228
ARIMA (2,2,1)	231.71	14.52631

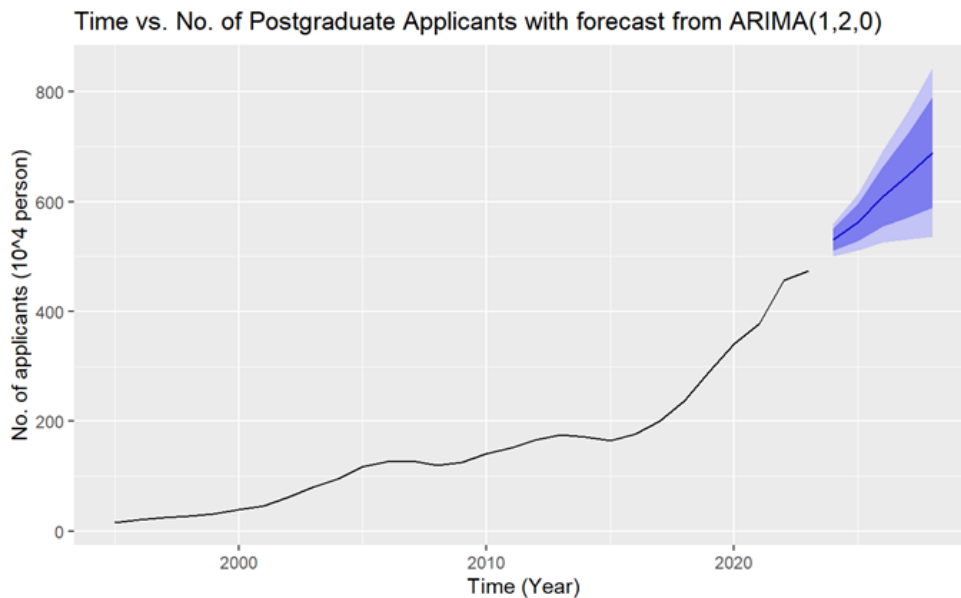
**Table 2.** Different ARIMA models for the enrollment data and their corresponding AICc and RMSE

Model	AICc	RMSE
ARIMA (3,2,0)	205.21	2.514218
ARIMA (2,2,0)	205.63	2.626658
ARIMA (1,2,0)	221.32	3.323556
ARIMA (3,2,1)	207.8	2.513542
ARIMA (2,2,1)	205.47	2.528858
ARIMA (1,2,1)	216.05	3.015222

### 3. Results and Discussion

#### 3.1. Applicant Data

Fig 8 and Table 3 records the predicted number of applicants in the next 5 years by the prediction of the ARIMA model. We can see from the figure that the number of applicants will continue to rise steadily.

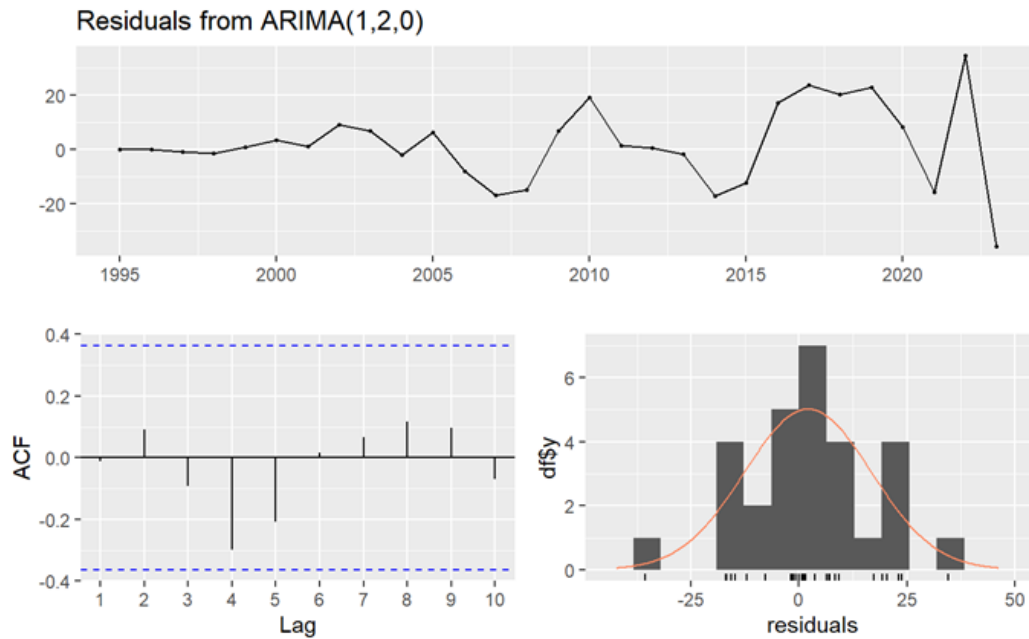


**Figure 8.** Number of postgraduate applicants in China (unit:  $10^4$  person) from 1995 to 2028 with forecast from 2024 to 2028 from ARIMA (1, 2, 0) included

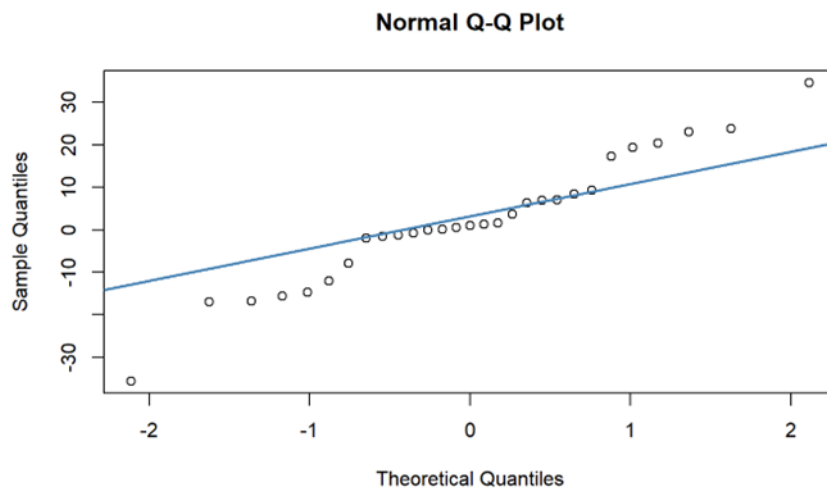
**Table 3.** Forecasted number of postgraduate applicants in China from 2024 to 2028 of ARIMA (1,2,0)

Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2024	530.1498	510.3260	549.9736	499.8319	560.4677
2025	561.9709	528.2094	595.7324	510.3372	613.6046
2026	608.9104	554.6414	663.1795	525.9131	691.9078
2027	646.4550	570.6747	722.2354	530.5590	762.3511
2028	689.8379	589.1809	590.4949	535.8963	843.7795

Checking the residues, we can see that the residues were small in earlier years, but gradually increased in more recent years, although the ACF plot in Fig 9 does not show an excess over critical value. The distribution of the residues is roughly normal from the histogram, although there is relatively high density than normal distribution near the 25% and 75% quartile. In the normal Q-Q plot of the residues in Fig 10, we can see that the residuals roughly follow a straight line, which confirms that the residues are overall normal and hence white noise. Lastly, the Ljung-Box test provides a p-value being 0.3674, greater than the critical 0.05, hence the residues are white noise.



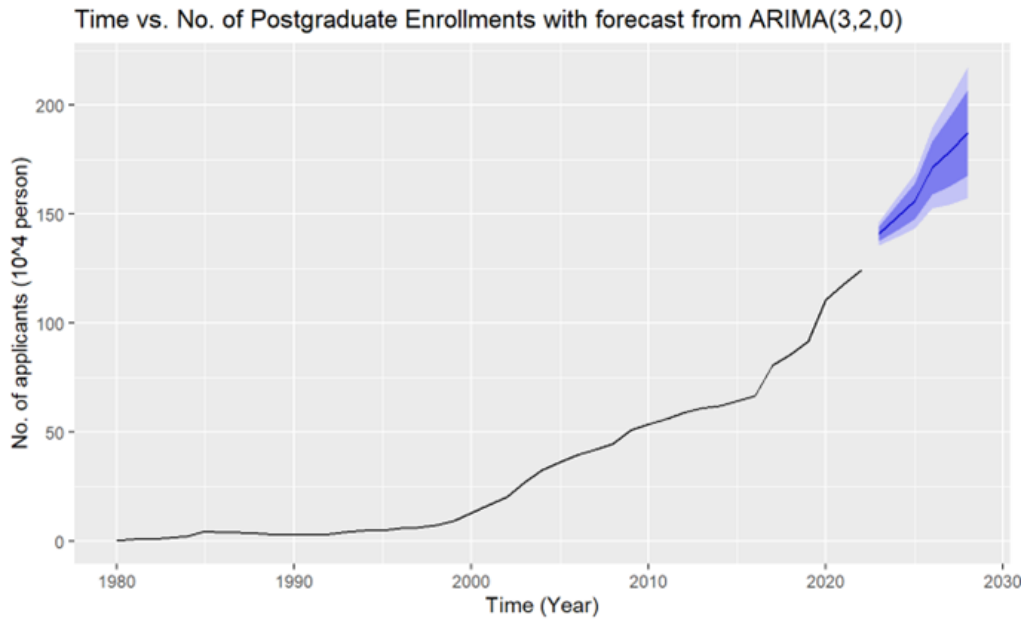
**Figure 9.** Residues of the model for the applicant data, with residues each year (up), ACF plot (down left), histogram (down right)



**Figure 10.** Normal Q-Q plot of the residues of the model for the applicant data

### 3.2. Enrollment Data

We follow a similar procedure for the enrollment data. Fig 11 and Table 4 records the predicted number of applicants in the next 6 years by the prediction of the ARIMA model. We can see from the figure that the number of enrollments will continue to rise steadily.

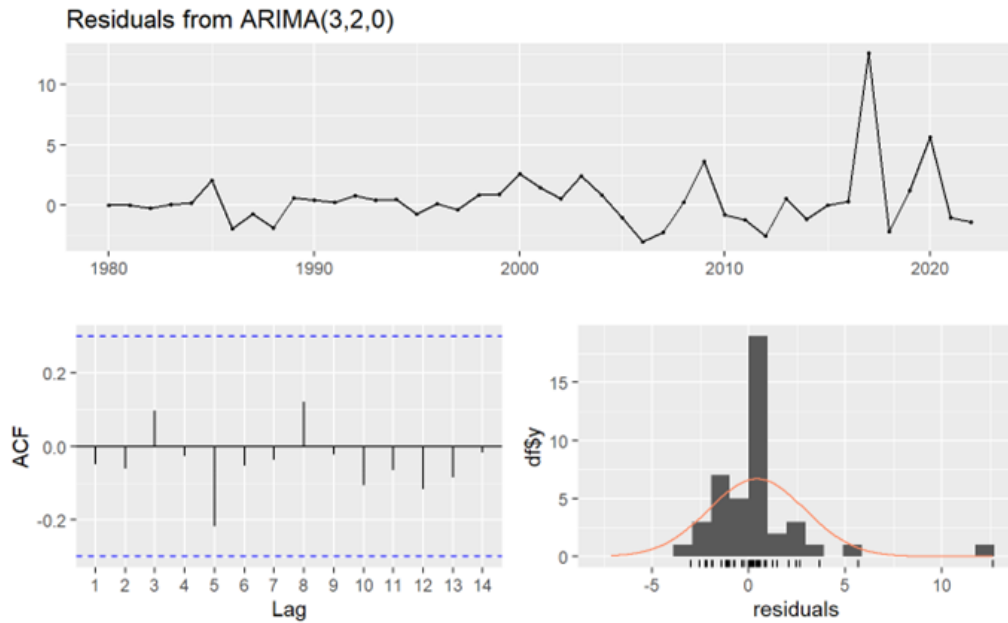


**Figure 11.** Number of postgraduate enrollments in China (unit:  $10^4$  person) from 1980 to 2028 with forecast from 2023 to 2028 from ARIMA (3, 2, 0) included

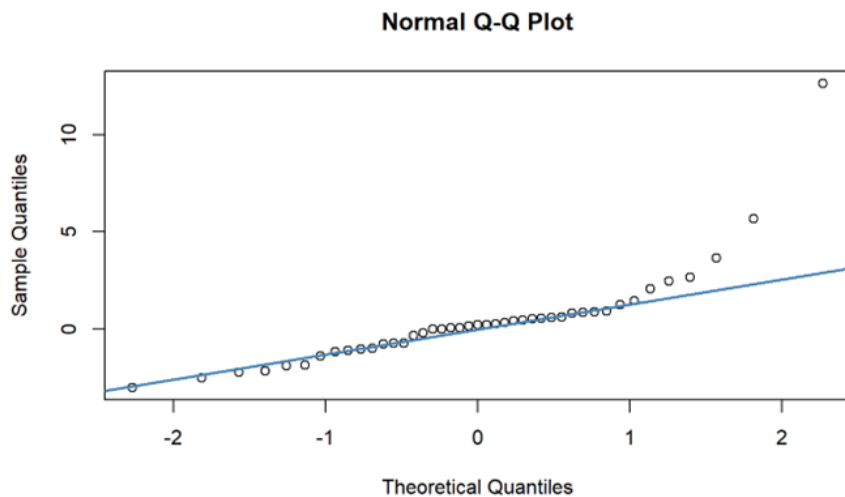
**Table 4.** Forecasted number of postgraduate enrollments in China from 2023 to 2028 from ARIMA (3,2,0)

Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2023	141.0319	137.6044	144.4595	135.7900	146.2739
2024	148.2410	142.3561	154.1258	139.2409	157.2410
2025	156.0066	147.8206	164.1927	143.4871	169.5262
2026	171.2968	159.1762	183.4175	152.7599	189.8337
2027	178.8539	162.8057	194.8021	154.4632	203.2446
2028	187.5191	167.8020	207.2363	157.3643	217.6739

Checking the residues, we can see that the residues are small in earlier years, and gradually became larger in later years, with the residue at 2017 being a huge anomaly. Even so, the ACF plot in Fig 12 does not show any excess over the critical value. The histogram shows that, besides the anomaly value that is clearly from 2017, the rest of the distribution decently fit the normal curve. The normal Q-Q plot in Fig 13 also shows a straight line, which also indicates a normal distribution of the residues. Lastly, the Ljung-box test passed with p-value being 0.6504, greater than the critical 0.05, hence we can conclude that the residues are indeed white noise.



**Figure 12.** Residues of the model for the enrollment data, with residues each year (up), ACF plot (down left), histogram (down right)



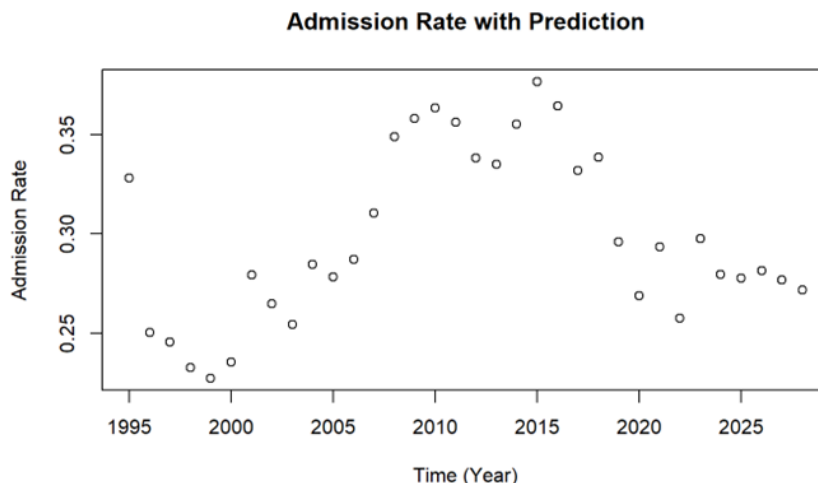
**Figure 13.** Normal Q-Q plot of the residues of the model for the enrollment data

**3.3. Combination of Both Data**

We can put the data together and obtain more observations, most notably about the admission rate, defined by dividing the enrollments by applicants.

$$Admission\ rate = \frac{Enrollment}{Applicant} \tag{1}$$

Fig 14 is the plot of the admission rate from 1995 to 2028, with the value from 2023 to 2028 (rightmost 6 points on the figure) being predicted values, and Table 5 being the numerical values. In the figure we can see that from the prediction, the admission rate is mostly stable around 27% to 28%.



**Figure 14.** Postgraduate admission rate from 1995 to 2028, with the values from 2023 to 2028 being predictions

**Table 5.** Postgraduate admission rate from 2018 to 2028, with the value of 2023 to 2028 being predictions

Year	Point Forecast
2018	33.86987%
2019	29.58503%
2020	26.87692%
2021	29.35149%
2022	25.75492%
2023	29.75357% (Predicted)
2024	27.96210% (Predicted)
2025	27.76062% (Predicted)
2026	28.13169% (Predicted)
2027	27.66688% (Predicted)
2028	27.18307% (Predicted)

#### 4. Conclusion

This paper has concluded using ARIMA models that the postgraduate applicants and enrollment will continue to rise in their different rates in the near future, while the postgraduate admission rate will be mostly stable.

ARIMA model itself is a powerful tool that has proven to be useful in the analysis of time series. This alone makes the research reliable. In addition, in comparison to previous researches, not only did this concluded a different ARIMA model, but it also combines with the enrollment to provide further information and possible suggestions. These are advances in comparison to previous publications.

This research also has its drawbacks. For one, the data is not precise to the one-digit, and the difference in the time of recording data may have a difference on precision and accuracy, and hence impact the research result. It is also very possible that there is a change point in the time series. However, the data size is too small to perform appropriate change point detection on the model. A Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is not applied for the same reason. The failure of applying these two powerful models makes this research deficit of precision. Moreover, there is a large residue in the enrollment forecast model in 2017, which could be due to the implementation of an important policy that included the employed graduate students into the postgraduate entrance exam. This paper has not dug deep into the reason behind the large residue, and the possible causal relationship just proposed could be important to study in further

researches. Lastly, this paper considered the admission rate as a division of enrollments over applicants, but the admission rate could be studied alone and could fit a non-ARIMA model as well. This could also be an inspiration for further study.

## References

- [1] Laura Tucker. Ten Good Reasons to Go to Grad School. Postgraduate enrollment report, 2022.
- [2] Xue Gexi. The Number of Postgraduate Entrance Applicants is Estimated to be 4.91 million, Why the Anxiety. Working paper, 2023.
- [3] Shi Yi. Prediction of the National Graduate Postgraduate Enrollment Based on R Language and ARIMA Model. *Computer Knowledge and Technology*, 2018, 14.
- [4] Sun Mengjie, Chen Baofeng, Wen Chunhui, Ren Jinzheng. Postgraduate Enrollment Modeling and Forecast based on ARIMA models. *Decision and Reference*, 2010.
- [5] Hu Yun. Postgraduate Applicants Forecast based on QPSO-BP. *Microcomputer Information*, 2010.
- [6] Li Fengliang, Yuan Bentao, Liu Huiqin. Medium to Long Term Prediction of the Size of Chinese Graduate Students in School: An International Comparative Perspective. *Higher Education Research*, 2008, 29 (5): 7.
- [7] Zhang Wei, Wang Jinsong. Prediction of Graduate Education Scale and Comparison between China and the United States. *Degree and Graduate Education*, 2022, 2: 1 - 7.
- [8] Liu Shucui, Yin Ping. Prediction of the Development Scale of Graduate Education Based on BP Neural Network. *Chinese Journal of Social Medicine*, 2008, 3: 3.
- [9] Yin Fayue. Research on the Current Situation of Graduate Employment and Prediction of Employment Prospects. *China Higher Education Evaluation*, 2007, 3: 5.
- [10] Li Hongxia, Li Chuanwei. Application of Artificial Neural Network Models in Predicting Graduate Enrollment Quantity. *Mathematical Practice and Understanding*, 2009, 12: 7.