

基于ARIMA模型的中国研究生招生人数预测分析

贺成婷, 把月佳*

昌吉学院, 数学与数据科学学院, 新疆 昌吉

收稿日期: 2024年6月29日; 录用日期: 2024年7月23日; 发布日期: 2024年8月23日

摘要

随着我国研究生招生规模的逐年增加, 国家宏观调控政策越来越需要适应社会和经济发展的需要, 因此, 编制年度研究生招生计划需要有理有据, 并构建编制研究生年度招生计划的合理模型。为了能更好地预测和确定我国研究生年度招生计划, 在借鉴一些学者和研究人员的成果的基础上, 本文采用时间序列分析方法对中国研究生招生人数进行预测, 为相关部门制定未来可持续发展计划提供参考, 随着越来越多的学生选择考研, 其原因也呈现出多样化趋势, “考研热”也已成为一种不容忽视的教育现象, 但是学术界对中国研究生招生人数研究则少有论及。因此, 本文首先对1984~2023年间中国研究生招生人数数据进行收集整理, 然后运用Eviews10软件对数据建立了最优模型ARIMA(0,1,3)。过程中用到了时序图法、差分运算法、单位根检验法、残差检验法等方法完成了对未来五年研究生招生人数的预测。最后, 研究表明, 2024年的考研报名人数相较于前几年出现了一定程度的下降。这一变化引发了人们对考研制度和教育发展的思考。有人认为, 这可能是因为就业形势改善, 一部分毕业生选择了直接就业而非继续深造。另一方面, 也有人指出, 这可能是对考研竞争激烈程度的一种回应, 一些考生认为自己的竞争力不足, 选择放弃考研, 转而寻找其他发展路径。这些结论和建议旨在帮助中国研究生招生发展更加稳健的同时也要保证质量方面的严格把关, 从而为科学快捷编制年度研究生招生计划提供科学理论依据。

关键词

ARIMA模型, 中国研究生, 预测分析

Predictive Analysis of Graduate Enrollment in China Based on ARIMA Model

Chengting He, Yuejia Ba*

School of Mathematics and Data Science, Changji University, Changji Xinjiang

Received: Jun. 29th, 2024; accepted: Jul. 23rd, 2024; published: Aug. 23rd, 2024

*通讯作者。

文章引用: 贺成婷, 把月佳. 基于 ARIMA 模型的中国研究生招生人数预测分析[J]. 现代管理, 2024, 14(8): 1906-1915.
DOI: 10.12677/mm.2024.148221

Abstract

With the increasing scale of graduate enrollment in our country year by year, the national macro-control policies are increasingly needed to adapt to the needs of social and economic development. Therefore, the preparation of annual graduate enrollment plans needs to be reasonable and well founded, and a reasonable model for the preparation of annual graduate enrollment plans needs to be constructed. In order to better predict and determine the annual enrollment plan for graduate students in China, based on the achievements of some scholars and researchers, this article uses time series analysis method to predict the number of graduate students enrolled in China, providing reference for relevant departments to formulate future sustainable development plans. With more and more students choosing to take the postgraduate entrance examination, the reasons for this have shown a diversified trend, and the “postgraduate entrance examination fever” has become an educational phenomenon that cannot be ignored. However, there is little research in the academic community on the number of Chinese graduate students admitted. Therefore, this article first collected and organized data on the number of Chinese graduate admissions from 1984 to 2023, and then used Eviews9 software to establish the optimal model ARIMA(0,1,3) for the data. During the process, methods such as time series diagram, differential operation algorithm, unit root test, and residual test were used to predict the number of graduate admissions for the next five years. Finally, research shows that the number of applicants for the postgraduate entrance examination in 2024 has decreased to a certain extent compared to previous years. This change has triggered people's reflection on the postgraduate entrance examination system and the development of education. Some people believe that this may be due to the improvement of the employment situation, with some graduates choosing direct employment rather than continuing their studies. On the other hand, some people have pointed out that this may be a response to the fierce competition in the postgraduate entrance examination. Some candidates believe that their competitiveness is insufficient and choose to give up the examination and instead seek other development paths. These conclusions and recommendations aim to help the development of graduate enrollment in China become more stable while ensuring strict quality control, thus providing a scientific theoretical basis for the scientific and efficient preparation of annual graduate enrollment plans.

Keywords

ARIMA Model, Chinese Graduate Students, Predictive Analysis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,中国硕士研究生的招收量持续扩大,伴随着扩招趋势,高校正承受着更多的压力,这使得他们必须提高并更新他们的研究生管理和服务水平。因此,如何在研究生规模不断扩大的背景下,提高教育水平和质量,还有待进一步的研究[1]。所以,对研究生招生规模进行科学、合理地预测,对于保障研究生教育的整体质量具有重要意义[2]。它能够教育单位把控教育资源的分配提供依据。马彩红等人借由制度变迁理论,运用制度变迁理论深挖自改革开放至今中国研究生招生体系的变化历程,揭示了这一过程中的内部逻辑并找出其中的关键特性与法则、探讨了可能出现的问题,对未来的招生体制改革趋势进行了合理的预测[3]。曹有容等人采用描述性统计及 GM(1,1)灰色模型对广西高等医学院校研究生培养整体情况及发展趋势进行了预测分析。得出随着广西高等医学院校研究生招生规模不断扩大,保持招

生规模稳步增长的同时进一步优化学科专业结构, 加强相关紧缺医学专业研究生的培养力度是研究生教育的重中之重[4]。谭军等人采用了 Logistic 预测模型, 并用遗传算法求得其待定系数, 求得了模拟值与实际招生数拟合最好的模型, 对湖北省过去二十年的硕士研究生录取情况进行了深入的分析和预测, 得出要想确保湖北省研究生教育的可持续增长, 要增强对于提升研究生教育价值的重要性意识, 积极改进其教学环境, 同时, 要转变关于中央政府、部门及省级单位举办研究生教育的传统思维, 强化地域教育管理机构的作用, 并对当地研究生教育的构成做出适当的调整, 减少过分强调行业学院或省政府所属大学的情况, 发挥本地研究生教育的整体优势[5]。汪瑾璇等人为了进一步提高招生管理效率, 通过研究高校研究生招生管理落实现状分析、工作中存在的问题、加强高校研究生招生工作管理的对策, 提出三点招生管理强化途径: (1) 提高招生管理人员综合素质; (2) 完善招生管理制度; (3) 优化招生管理的培训指导[6]。李红霞等人使用了人工神经网络模型来解决黑龙江省的研究生招生历史数据中存在的信息分散和众多潜在的影响因素的问题, 研究中利用这个模型对每年的人数变化进行了深入的探索和系统的分析, 并建立了一个有效的预测模型成功地预估了未来的三年内可能出现的招生人数的变化趋势[7]。许多要素会影响某一区域的研究生招生数量, 如国家的教育方针、地方的教育法律和当地的社会经济状况等等。所以, 用一种方法来预估特定区域的研究生招生量是不可靠的。为了通过合理的预测来控制研究生的招生规模, 以此作为调整教育资源的重要参考依据, 有助于确保整个研究生教育的质量得到有效保护, 我们考虑到 ARIMA 模型对于短期预测的高精确度, 本篇论文将其应用到了研究生教育规模的预测中, 并通过中国历年的研究生招生数据作为实际案例进行验证。基于预测的结果, 我们进一步探讨了中国的研究生招生数目发展轨迹及当前研究生教育的情况, 从而提出了一系列有助于实现研究生教育规模与品质平衡提升的相关意见。

自改革开放开始至今, 硕士生录取政策已根据政治、经济状况变化和国民深层改革进程持续调整。近些年, 由于硕士生更积极地制定未来计划, 我们认为“所有历史都与现在息息相关”, 过去的经验塑造着当前乃至未来的走向。基于此, 本文利用时间序列模型精确预测过往数据并以此作为基准, 对 1984 至 2023 年的中国研究生招录规模展开预测分析, 旨在揭示其历史演进过程, 探索发展的法则, 发现潜在问题, 同时考虑目前的形势和需求, 推测研究生录取政策的发展趋势和路线, 并给出有实际指导意义的改良方案, 为我们优化中国的研究生录取制度提供了有益的参考意见。

2. ARIMA 模型介绍

我们都知道, 在获得观察值序列后首先我们要对其进行平稳性检验, 当观察值序列是平稳序列时我们才可以对其进行下一步的预测分析。但是实际生活中绝大部分序列都是非平稳的(如本文数据得到的原始序列), 所以我们一般会对非平稳序列采取差分运算使其显示出平稳序列的性质。而对差分后的平稳序列我们一般可以使用求和自回归移动平均模型来进行拟合预测, 简记为 ARIMA(p,d,q)模型:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (1)$$

(1)式中, $\nabla^d = (1-B)^d$; $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, 为平稳可逆 ARMA(p,q)模型的自回归系数多项式, $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为平稳可逆 ARMA(p,q)模型的移动平均系数多项式。

求和自回归移动平均模型的来由: d 阶差分后的序列:

$$\nabla^d x_t = \sum_{i=0}^d (-1)^i C_d^i x_{t-i} \quad (2)$$

(2)式中, $C_d^i = \frac{d!}{i!(d-i)!}$, 表示差分后序列等于原序列的若干序列值的加权和, 对差分序列又可以拟合自回归移动平均 ARMA 模型, 这就是求和自回归移动平均模型的来由。简记为:

$$\nabla^d x_i = \frac{\Theta(B)}{\Phi(B)} \varepsilon_i \quad (3)$$

(3)式中, $\{\varepsilon_i\}$ 为零均值白噪声序列。ARIMA 本质上是结合了差分运算和 ARMA 模型, 而 ARMA 模型的分析方法已经十分成熟, 因此, 我们运用适当的差分运算对非平稳的序列进行处理后再进行的拟合分析(ARIMA 模型)也是十分简便可靠的。

3. 实证分析与结果

为了预测我国未来五年的研究生招生人数, 我们采用了 1984 年至 2023 年的相关数据(数据均来源于《中国统计年鉴》), 并使用 Eviews10 统计分析软件建立了一个合适的模型。

从 1984 年至 2023 年间的中国研究生招生数据时序图(见图 1)可以观察到, 整体上, 中国研究生招生人数呈现出指数级增长, 初步判断该时间序列并不稳定。

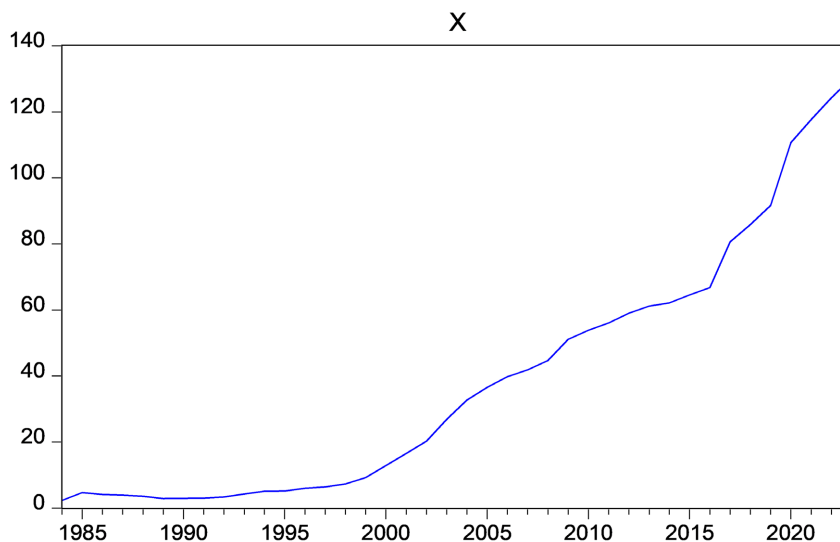


Figure 1. Time series chart of China's graduate enrollment data from 1984 to 2023

图 1. 1984~2023 年中国研究生招生人数数据的时序图

1) 自相关图检验法

中国的研究生人数序列的自相关系数不是很快地趋于零, 即缓慢下降, 再次判断该序列不是一个平稳序列(如图 2)。

2) 单位根检验

通过以上分析该序列很明显不是平稳序列, 我们又对其做了单位根检验(含常数项)结果仍然显示它不是一个平稳序列(见图 3)。

接下来, 我们将对原序列进行一阶差分平稳化处理, 再对差分后的序列进行检验。本文通过 eviews10 软件对该序列进行一阶差分处理。由图 4 可以看出一阶差分后的序列似乎总是在一个常数附近波动, 但是在后面几年有较大波动, 根据之前的经验, 一般来说一阶差分后的序列大多就平稳了, 而且, 图示法一般具有较大主观影响, 所以本文进一步对一阶差分后的数据进行更为详细的单位根检验来确定其平稳性。

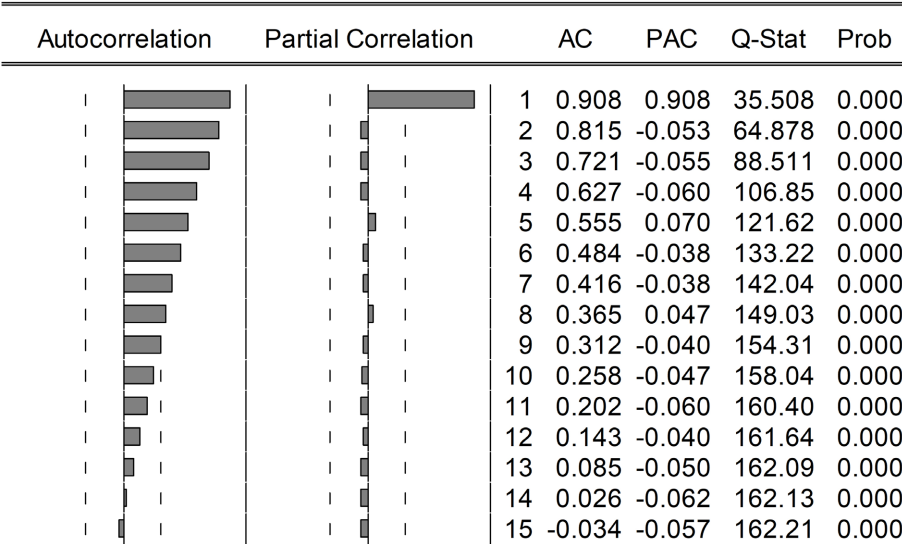


Figure 2. Autocorrelation chart of graduate enrollment in China from 1984 to 2023
图 2. 1984~2023 年中国研究生招生人数的自相关图

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	4.789658	1.0000
Test critical values:		
1% level	-3.610453	
5% level	-2.938987	
10% level	-2.607932	

Figure 3. Unit root test (0th order, including constant terms)
图 3. 单位根检验(0 阶、含常数项)

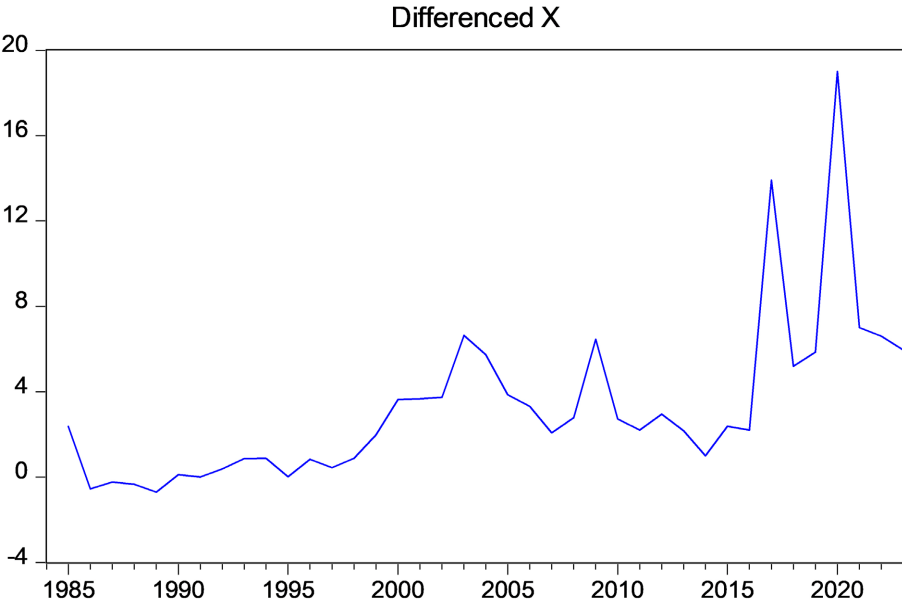


Figure 4. Time series after first-order differencing
图 4. 一阶差分后的时序图

如图 5 所示, 一阶差分后, 从实验结果得出: 1 阶、含常数项和趋势项形式下的 P 小于 0.05, 说明一阶差分后的序列是平稳的时间序列。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.511746	0.0003
Test critical values: 1% level	-4.219126	
5% level	-3.533083	
10% level	-3.198312	

Figure 5. Unit root test (1st order, including constant and trend terms)

图 5. 单位根检验(1 阶、含常数项和趋势项)

由图 5 直观的展现出, P 值等于 $0.0003 < 0.05$, t 统计量值为 -5.511746 , 这个值比 1%、5%、10% 显著水平下的都小, 拒绝原假设, 时间序列平稳。

单位根检验过后, 进一步对经过一阶差分处理后的中国研究生招生数据进行了纯随机检验得到图 6。检验结果显示各阶延迟下的 P 值都小于显著性水平 0.05, 所以可以得到一阶差分后的序列为一个平稳的非白噪声序列, 进一步证明了此时序列的平稳性。

































Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.474	0.474	9.4553	0.002
		2	0.403	0.230	16.485	0.000
		3	0.523	0.364	28.648	0.000
		4	0.171	-0.295	29.981	0.000
		5	0.133	-0.068	30.815	0.000
		6	0.054	-0.207	30.955	0.000
		7	-0.004	0.160	30.955	0.000
		8	0.075	0.153	31.243	0.000
		9	-0.004	0.045	31.244	0.000
		10	-0.028	-0.156	31.287	0.001
		11	0.054	-0.023	31.456	0.001
		12	-0.024	-0.063	31.489	0.002
		13	-0.011	0.117	31.497	0.003
		14	0.030	0.045	31.553	0.005
		15	0.003	0.053	31.554	0.007
		16	0.068	0.002	31.878	0.010

Figure 6. Autocorrelation diagram after first-order differencing

图 6. 一阶差分后的自相关图

3) 模型的识别与定阶

由一阶差分后的自相关图(图 6)显示, 延迟 1 阶、2 阶、3 阶的自相关系数大于两倍的标准误差。偏自相关图显示延迟 1 阶和 3 阶的偏自相关系数大于两倍标准差。根据 ACF 和 PACF 的特性选择建立模型, 建立 ARIMA(0,1,3)模型, 见图 7。

综合图 7 中的数据信息, 本文最终选择 ARIMA(0,1,3)作为 1984~2023 年中国研究生招生人数的预测模型, 此时模型是一阶、含常数项和趋势项的形式, 此时的单位根检验的 P 值也更小, 进一步说明了模型的可靠性。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.454969	1.126666	3.066544	0.0042
MA(1)	0.339583	0.055380	6.131856	0.0000
MA(2)	0.288386	0.055634	5.183641	0.0000
MA(3)	0.938759	0.034009	27.60340	0.0000
R-squared	0.541931	Mean dependent var		3.279023
Adjusted R-squared	0.502668	S.D. dependent var		3.880195
S.E. of regression	2.736381	Akaike info criterion		4.948064
Sum squared resid	262.0723	Schwarz criterion		5.118686
Log likelihood	-92.48725	Hannan-Quinn criter.		5.009282
F-statistic	13.80259	Durbin-Watson stat		1.907234
Prob(F-statistic)	0.000004			
Inverted MA Roots	.33+.91i	.33-.91i	-1.00	

Figure 7. Parameter estimation diagram of model ARIMA(0,1,3)

图 7. 模型 ARIMA(0,1,3)的参数估计图

4) 模型检验

确定了匹配模型后，此阶段会对已创建的模型实施重要性的评估，其重点在于检查模型对于数据抽取的完整度、参数的重要性测试及模型架构的最优化程度。检测的结果以图表形式展示于如(图 8)。可以从结果看出 ACF 和 PACF 都没有显著异于零，Q 统计量的 P 值几乎都大于 0.05，因此可以认为残差序列为白噪声序列，模型信息提取比较充分，整个模型比较精简，模型较优。

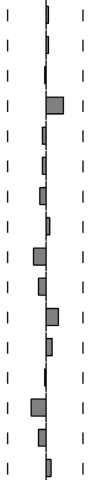
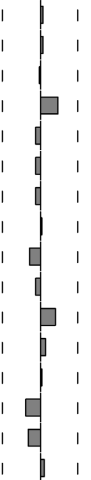
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.027	0.027	0.0317	
		2 0.031	0.031	0.0741	
		3 -0.006	-0.008	0.0757	
		4 0.154	0.154	1.1586	0.282
		5 -0.025	-0.034	1.1890	0.552
		6 -0.030	-0.038	1.2318	0.745
		7 -0.045	-0.040	1.3335	0.856
		8 0.034	0.015	1.3940	0.925
		9 -0.100	-0.093	1.9236	0.927
		10 -0.052	-0.040	2.0742	0.956
		11 0.113	0.138	2.8054	0.946
		12 0.056	0.041	2.9940	0.965
		13 -0.003	0.013	2.9945	0.982
		14 -0.116	-0.115	3.8573	0.974
		15 -0.056	-0.100	4.0691	0.982
		16 0.046	0.038	4.2136	0.989

Figure 8. Residual test of model ARIMA(0,1,3)

图 8. 模型 ARIMA(0,1,3)的残差检验

从模型的拟合图(图 9)可以进一步看出该模型的拟合效果较好可以用于对中国研究生的招生人数进行下一步预测分析。

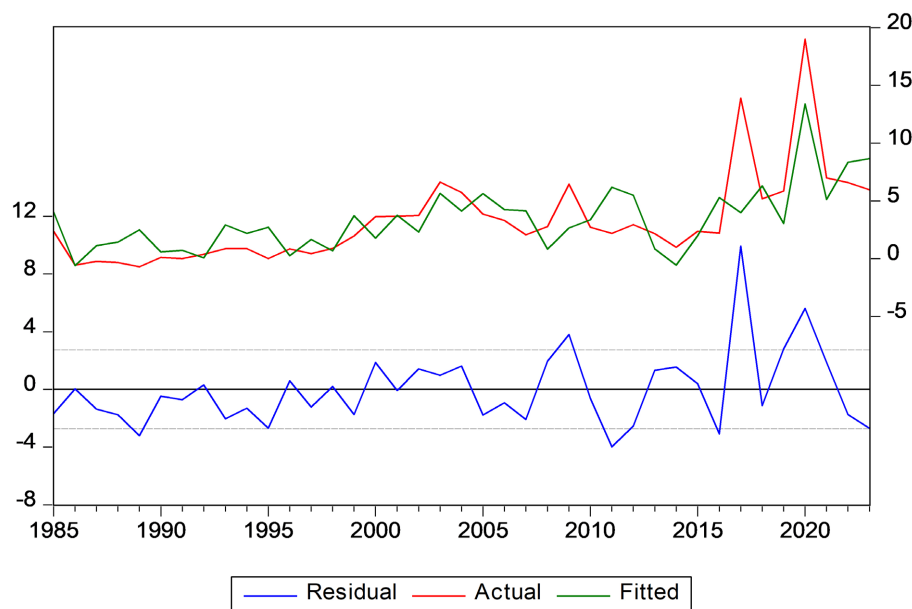


Figure 9. The fitting effect of the model ARIMA(0,1,3)

图 9. 模型 ARIMA(0,1,3) 的拟合效果图

5) 招生人数预测

首先, 本文对 1984~2023 年研究生招生人数样本内预测图做分析, 发现拟合效果较好, 如图 10 所示。

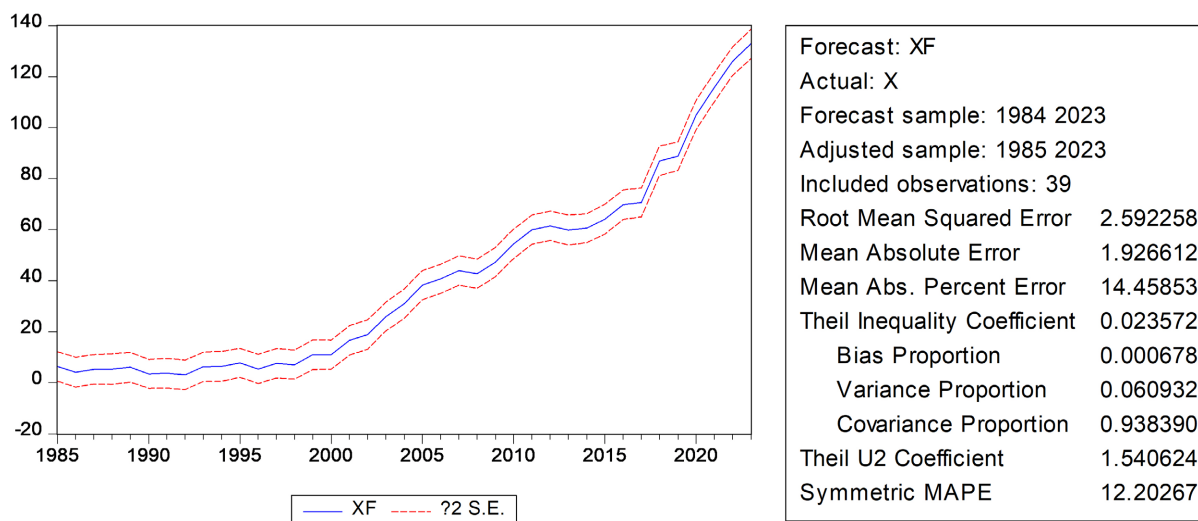


Figure 10. Sample level prediction of graduate enrollment from 1984 to 2023

图 10. 1984~2023 年研究生招生人数样本内预测图

如图 10 所示, 我们得到该模型的偏差率和方差比率占比较小, 协方差比率占比很大而且三者的和等于 1。可以看出该预测模型较好, 可以进一步对数据进行分析预测。

最后, 将序列的真实值和预测值组成一个数组(图 11), 做出两者的时序图对其进行比较可以看出实际值序列和预测模型序列的拟合程度很好, 预测值与真实值的差异较小。

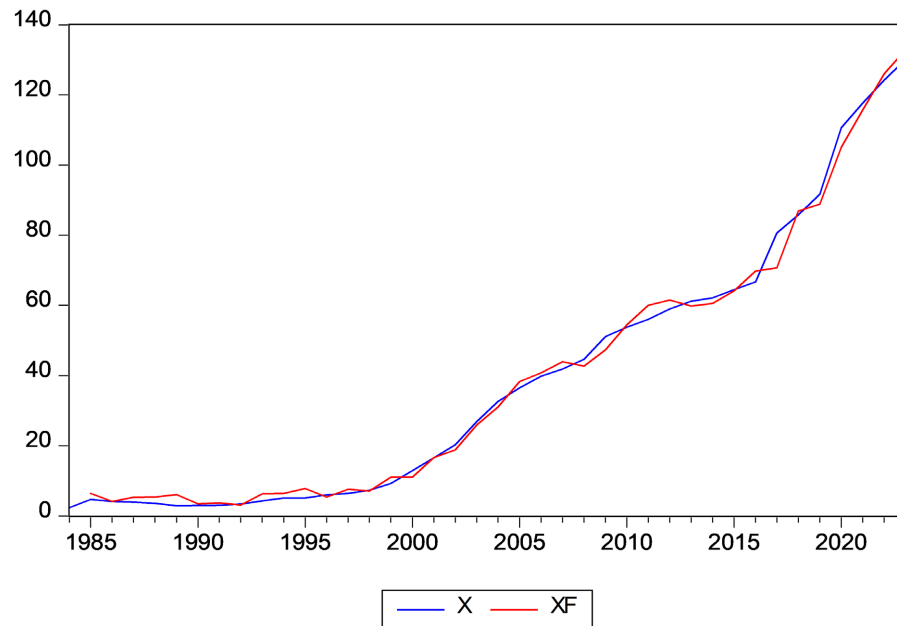


Figure 11. Fit between actual values and prediction models
图 11. 实际值与预测模型拟合图

6) 预测结果分析

由预测结果和原值的比较结果可以看出模型预测结果较好(见表 1), 相对误差较小, 成功预测了我国未来五年的研究生招生人数(见表 2)。未来 5 年我国研究生招生人数的上涨速度将继续增大, 我国相关教育部门将迎来新的挑战。

Table 1. Comparison of predicted and actual values
表 1. 预测值与实际值的结果比较

年份	2020	2021	2022	2023
实际值	110.655	117.652	124.2479	130.2
预测值	105.050	115.774	125.998	132.911

Table 2. Forecast of graduate enrollment in China in the next five years
表 2. 未来五年中国研究生招生人数预测值

年份	2024	2025	2026	2027	2028
预测值	130.8577	134.1466	137.43552	140.72440	144.01327

4. 结论

随着我国研究生招生规模的不断扩招, 研究生的招生质量和研究生的教育培养方面也迎来了巨大的挑战。本文运用统计学与时间序列的理论方法作为基础, 结合研究生招生的实际情况, 运用时间序列方法 ARIMA 模型对近几年研究生人数做预测研究, 得到结论如下:

依据 1984~2023 年间中国研究生招生人数数据, 利用 Eviews 软件对数据平稳分析处理, 取差分使得

数据成为平稳序列,然后根据 ACF 和 PACF 进行模型识别,通过模型检验与优化,最终建立 ARIMA(0,1,3)模型,预测 2024~2028 年未来 5 年的研究生招生人数,可以较好的反应研究生人数的趋势,但是该模型是在不考虑就业等的影响因素的情况下对未来趋势做预测,由于研究生人数受各种外界情况的影响,例如,如果研究生就业形势不太好的情况下,会导致研究生人数下降,此时预测误差会较大,所以该模型仅用来进行短期预测,如果模型被用来预测长期研究生人数,就需要进一步加强对模型的研究,提高模型预测的精度,使得预测更为精准,预测更接近实际值。从总体上看,在过去的几年里,考研报名人数连续增长,主要受到多方面因素的影响。一方面,随着高等教育的普及,本科生数量增加,加上经济下行的影响,许多考生认为继续深造可以提高自己的就业竞争力。另一方面,一些考生希望通过考研来提升自己的学校层次,以应对就业压力。然而,随着考研热度的上升,考研的难度也在不断增加。越来越多的考生意识到,花费大量的时间和经济成本去获取一个硕士学位,并不一定能够带来显著的就业优势。特别是在经济不景气的情况下,许多考生开始重新评估考研的性价比,认为与其花费大量的时间和金钱去考研,不如及早进入职场积累工作经验。此外,研究生培养的质量把控越来越严格,考试难度增加,使得“混”毕业变得越来越难。一些考生在权衡利弊后,可能会选择不考研或者推迟考研计划。这也提醒了广大考生,考研有风险,读研需谨慎。在做出决定前,应充分考虑自己的研究兴趣、能力以及未来的职业规划。总的来说,2024 年考研报名人数的下降反映了考生对考研的理性看待,以及对研究生教育投资回报的重新评估。这也预示着“考研热”可能开始降温,考生和家长会更加理性地考虑是否继续深造。

参考文献

- [1] 边国翠, 许晓兵. 基于 GM(1,1)模型对近几年全国研究生招生人数进行分析与预测[J]. 物流工程与管理, 2017, 39(2): 209-210.
- [2] 孙梦洁, 陈宝峰, 温春卉, 等. 基于 ARIMA 模型的研究生招生规模建模与预测[J]. 统计与决策, 2010(12): 60-62.
- [3] 马彩红. 基于制度变迁理论的我国硕士研究生招生制度研究(1978-2020) [D]: [硕士学位论文]. 长沙: 湖南师范大学, 2022.
- [4] 曹有容, 梁文杰, 白欣玉, 等. 基于灰色模型的广西高等医学院校研究生培养规模预测分析[J]. 广西医科大学学报, 2022, 39(12): 2026-2030.
- [5] 谭军, 沈成武. 基于遗传算法的硕士生招生数 Logistic 预测模型[J]. 武汉交通科技大学学报, 1999(4): 399-402.
- [6] 汪瑾璇. 强化高校研究生招生工作管理的对策分析[J]. 产业与科技论坛, 2023, 22(24): 286-288.
- [7] 李红霞, 李传威. 人工神经网络模型在研究生招生数量预测中的应用[J]. 数学的实践与认识, 2009, 39(12): 27-33.