Statistics and Application 统计学与应用, 2026, 15(4), 155-162
Published Online April 2026 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2026.154079

Hans汉斯

基于ARIMA对沪深300指数分析

余飞飞，郑陈轩*

荆楚理工学院数理学院，湖北 荆门

收稿日期：2026年3月15日；录用日期：2026年4月5日；发布日期：2026年4月21日

摘　要

近年来，我国资本市场发展迅速，股市越来越受人们关注，沪深300指数作为投资收益率标杆，有利于投资者对市场投资回报率分析。因此，本文将时间序列ARIMA模型对沪深300指数收盘价短期预测。经实验证明，将预测收盘价和实际收盘价比较，模型短期预测效果较佳。

关键词

沪深300指数，时间序列，ARIMA模型，预测

Analysis of the CSI 300 Index Based on ARIMA

Feifei Yu, Chenxuan Zheng*

School of Mathematics and Physics, Jingchu University of Technology, Jingmen Hubei

Received: March 15, 2026; accepted: April 5, 2026; published: April 21, 2026

Abstract

In recent years, China's capital market has developed rapidly, and the stock market has received increasing attention. The CSI 300 Index, as a benchmark for investment returns, is beneficial for investors to analyze market investment return rates. Therefore, this paper will use the time series ARIMA model to make short-term forecasts of the closing prices of the CSI 300 Index. Experiments have shown that by comparing the predicted closing prices with the actual closing prices, the model performs well in short-term forecasting.

*通讯作者。

文章引用: 余飞飞, 郑陈轩. 基于 ARIMA 对沪深 300 指数分析[J]. 统计学与应用, 2026, 15(4): 155-162.
DOI: 10.12677/sa.2026.154079

Keywords

CSI 300 Index, Time Series, ARIMA Model, Forecasting

1. 绪论

1.1. 研究背景和意义

中国股票市场成立有 30 多年载了，自 1986 年，我国第一次发行股票，到上交所、深交所正式成立，再到如今，创业板由核准制转变成注册制，我国的股票市场发生了巨大的变化，不断走向成熟。机构投资者、个人投资者也从股市快速发展不断成熟，对股市有自己独立思考和判断。

相比国外的发达国家，我国股票市场还有很多不完善的方面，我国股票波动比较剧烈，股价会受一定的技术分析层面影响。人们发现沪深 300 指数的波动性，可以很好的反应中国股市的变化，虽然沪深 300 没有上证指数成立的早，但是沪深 300 的收益率，被很多投资者机构、和投资组合作为标杆。沪深 300 是指上海和深圳证券市场选取规模大、流动性好的 300 只股票，近年来沪深 300 指数，受到很多投资者去追捧。因此，对沪深 300 指数研究预测意义重大，也有多学者和机构研究人员正在研究该指数。分析和研究结果可以得出投资的建议，同时这也可以给国家实施宏观政策给予重要的理论建议。

1.2. 文献综述

1.2.1. 不同方法对沪深 300 指数的研究状况

(1) 关于机器学习对沪深 300 指数的研究

冯宇旭、李裕梅在 2019 年对沪深 300 指数预测，使用了 SVR 模型、和 Adaboost 模型和 LSTM 模型，同时用 SVR、Adaboost 和 LSTM 进行岭回归集成，这几个模型预测效果综合比较，LSTM 的效果最好[1]。孙武彪在 2019 年使用 BP 神经网络模型预测沪深 300 指数日收盘价，单一的 BP 神经网络模型拟合效果较差，通过了遗传算法优化 BP 神经网络模型拟合效果良好[2]。

(2) 关于时间序列模型对沪深 300 指数

李冰在 2016 年使用 ARIMA 模型和人工神经网络模型预测沪深 300 指数日收盘价，研究结果表明，ARIMA 模型短期预测效果更优相比 BP 神经网络模型，但是在不考虑其它因数时 ARIMA-ANN 组合模型拟合效果最佳[3]。朱文秀在 2020 年研究沪深 300 指数日收盘价数据，提出一种神经网络的 LSTM 模型与时间序列 ARIMA 模型集成模型来预测沪深 300 指数日收盘价，该集成模型拟合效果最优，ARIMA 模型短期拟合效果也很好，都比 LSTM 模型预测效果好[4]。徐鑫在 2015 年提出 Copula-ARIMA-GJR-GARCH 模型，用该模型估计沪深 300 指数和上证综合指数的联合分布，并通过对两种股票指数的未来收益率进行预测，发现该模型可以容易发现股票指数之间的条件相关性[5]。谢太峰、王硕、苏磊在 2017 年使用 ARMA-GARCH 模型研究监管政策对沪深 300 的收益率波动的影响[6]。研究表明监管政策对股指波动影响不显著。

由上可得，现今有很多学者研究沪深 300 指数研究，通过文献分析总结，在沪深 300 指数收盘价时，

单一的机器学习和神经网络模型拟合效果不佳，如果使用神经网络模型和时间序列模型集成拟合效果最优，但是只考虑收盘价历史因素，ARIMA 模型短期拟合效果很好。

1.2.2. 有关时间序列模型对股市研究状况

胡静在 2013 年对我国股市预测使用 ARIMA-NN 模型与 GARCH 族模型的比较研究，得出组合模型的拟合效果最佳[7]。张颖超在 2019 年对上证指数分析与预测数据研究，使用 ARIMA 模型短期预测，拟合效果较好，误差率小于 10%，第一天的预测精确度非常高[8]。万建强、文洲在 2001 年对比 ARIMA 模型与 ARCH 模型预测股指，研究表明 ARIMA 模型在短期预测股价的波动效果良好，但是在不同期间 ARIMA、ARCH 模型拟合效果不同[9]。

由上可得，很多学者使用 ARIMA 模型在股市中做研究，可总结出 ARIMA 模型在股市中短期预测中拟合效果良好。

综上所述，可使用 ARIMA 模型短期预测沪深 300 指数收盘价。

2. 差分自回归移动平均模型

2.1. 平稳时间序列

平稳时间序列可以理解为序列的统计性质不随着时间而发生改变，一般时间序列的分析都需要平稳。只有平稳了，差分自回归移动平均应用才是有效的。

可以根据时序图和相关图主观地判断该序列是否平稳，但是这种判断不具有说服性，为了平稳性客观的检验，也可以通过 DF 检验、ADF 检验的单位根检验方法。

2.2. 差分自回归移动平均模型

通过差分后平稳的时间序列，可以使用 ARIMA (p, d, q)模型进行拟合，该模型实际是由 ARMA 模型和差分运算的结合。而 ARMA 模型是由 AR (p)自回归模型和 MA (q)移动平均模型结合。差分后时间序列要通过白噪声检验，确定该序列不是白噪声序列，才能进一步地预测。中心化的 ARMA (p, q)模型数学表达式：

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (\varphi_0 = 0) \tag{3-1}$$

由于 3-1 公式易知，ARIMA 模型原理是解决线性问题，它受一定非线性的因素影响，所以只能在短期拟合效果才比较好。

以下简述 ARIMA 模型建立步骤：

第一，时间序列平稳性检验。可以通过时序图、相关图主观判断，也可以通过单位根检验 ADF 检验客观判断。

第二，平稳性处理。如果序列不平稳，则可通过直接差分或对数差分，可以定下平稳化的阶数 d，使序列到达平稳(通过第一步检验差分后的序列是否平稳)；如果序列是平稳，可跳过第二步。

第三，确定 p 和 q 的阶数。可以通过自相关图和偏自相关图大概判断，也可以使用自动拟合判断，同时使用 AIC 准则，确定 p 和 q 的值，选取相对拟合最优模型。

第四，模型检验。确定参数的模型拟合后，要对残差序列进行白噪声检验，如果残差序列不是为白噪声序列，跳回第三步，相反，可以说明该模型拟合效果较佳。其次对模型的参数显著性检验，定阶的参数显著模型预测才更加精准。

第五，模型预测。使用建立好的模型进一步预测，把预测值和实际值进行对比。

3. 实例分析

3.1. 实证数据集

3.1.1. 数据的选取

通过网易财经网，下载沪深 300 指数日的数据，对比沪深 300 指数的几个基本指标，选取投资者较关心的收盘价指标作为研究对象，经过 EXCEL 简单的整理得到 2019 年 1 月 2 日到 2021 年 2 月 26 日的我国股市开市日的沪深 300 指数日收盘价的 522 条数据作为研究的数据集。然后将 522 条数据中选取 2021 年的 1、2 月份 35 条数据作为测试集，剩下的 487 条数据作为训练集，通过模型训练训练集得出 35 个收盘价的预测值，然后将其预测值与真实值进行比较，对模型拟合效果进行分析与评价。

之所以选择 35 个交易日预测，是因为中国股市发展较晚，股票市场各种制度不够完善，拥有大量的散户，具有很多投资者做短线投资。投资者和政府机构对短期股市涨跌比较关心，考虑到短期预测模型的适用性，虽然 ARIMA 模型受限于线性拟合，但是使用该模型在短期拟合效果良好。本文主要使用 R 软件进行处理、分析数据。

3.1.2. 数据的描述

对沪深 300 指数收盘价整体数据的走势做出时序图，对其走势整体的把握。时序图 1 所示：

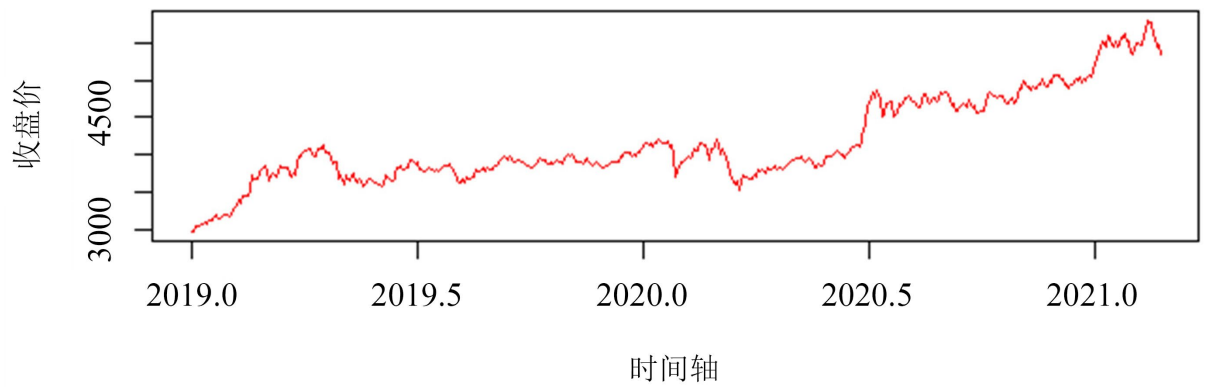

Figure 1. Time series chart of the daily closing prices of the overall CSI 300 Index
图 1. 整体沪深 300 指数日收盘价时序图

通过时序图，明显可以看出沪深 300 指数收盘价走势有一定的周期性，可以总结为六阶段，缓慢上涨、急速下跌、横盘震荡、缓慢上涨、急速上涨。在 2020 年的 3 月急速下跌，受全球疫情快速扩张影响，同时在当年的 6 月份急速上涨，受到美联储的大量放水起作用，可以看出沪深 300 收盘价受多因素影响[10]。

对整体的数据计算简单的描述性统计量，统计量如表 1 所示：

Table 1. Descriptive statistics of daily closing prices of the CSI 300 index
表 1. 沪深 300 指数日收盘价描述性统计量

mean	sd	max	min
4164.56	591.19	5807.72	2964.84

从标准差和极差统计量可以看出，沪深 300 指数日收盘价波动比较剧烈。

3.2. ARIMA 模型建立

在 70 年代 Box 和 Jenkins 提出的时间序列预测方法，其中自回归移动平均模型，简称：ARIMA 模型。它内容包含 AR、MA、ARMA 模型具体内容。目前 ARIMA 模型已经广泛的应用于多个邻域。

3.2.1. 序列平稳性检验

从图 1 沪深 300 指数日收盘价时序图可以看出，整体数据有明显的波动趋势，可以初始判定该序列是不平稳的，同时可以从 ACF 自相关图和 PACF 偏自相关图也可以看该序列是不平稳的。

由于以上的检验都是通过看图判断出序列不平稳，有一定的主观层面，可以通过 ADF 检验方法客观判断，对该序列做单位根检验，在 95%的显著水平下，设原假设为该时间序列是平稳。ADF 检验的结果 P 值都远大于 0.05，则拒绝原假设，说明该序列存在单位根，可以判定该时间序列为不平稳，判断结果与初始的图检验判断一致。

3.2.2. 序列平稳性处理

由于沪深 300 收盘价序列不平稳，通过直接差分法对该序列进行处理。差分运算实际是对整体序列信息的加工，所以每次差分都有一些信息丢失，避免过差分现象出现，要选择合适的阶数差分。该序列有一定的周期性，所以先对该序列进行一阶差分，然后对差分后序列再一次平稳性检验。

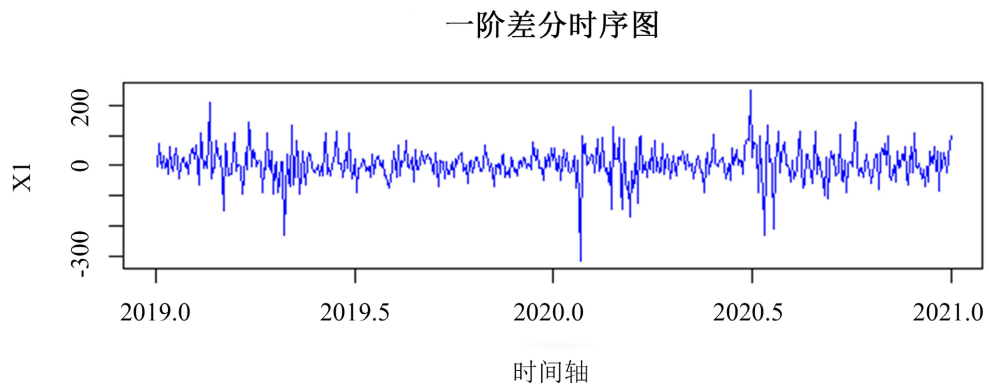

Figure 2. First-order difference of the daily closing price time series of the CSI 300 Index
图 2. 一阶差分沪深 300 指数日收盘价时序图

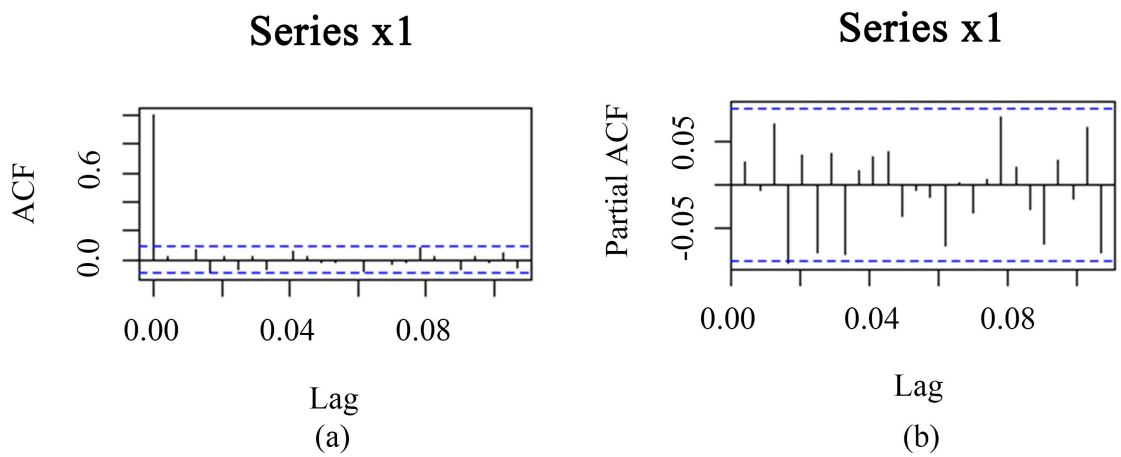

Figure 3. After first-order differencing: (a) ACF, (b) PACF plots
图 3. 一阶差分后：(a) ACF，(b) PACF 图

通过一阶差分时序图可以看出，该序列波动幅度比较小，并且基本平均分布在 0 轴的两侧；同时一阶差分后的 ACF 和 PACF 相关性检验出现了拖尾和截尾的现象，初步可以判断该序列是平稳的。其次对一阶差分序列进行 ADF 检验，尝试多种参数检验的 P 值结果都为 0.01，在 95%的显著水平下，可以接受原假设，说明该序列不存在着单位根。可以判定一阶差分的时间序列为平稳序列。

至此，对收盘价数据进行了基本的分析，下面对该序列进行模型的构建与具体分析。

3.2.3. 模型定阶

由于序列通过一阶差分后达到平稳，可以确定 d 为 1。然后确定 q 和 p 值，可以通过一阶差分后的 ACF 和 PACF 图，可以看出 ACF 截尾，PACF 也存在截尾，可以大致确定 p 为 2，q 也为 2，同时使用自动 ARIMA 拟合，的出 ARIMA (1，1，2)，经过对阶数的多次尝试和根据 AIC 最小准则，同时沪深 300 指数收盘价不受季节因素的影响，所以最终选择 ARIMA (1，1，2)对该序列进行预测。

3.2.4. 模型检验

(1) 对 ARIMA (1，1，2)模型的参数显著性检验。使用 auto.arima 拟合该序列，得出系数参数见表 2：

Table 2. Significance test of model parameters
表 2. 模型参数显著性检验

Variable	Coefficient	Sd.error	T_statistic	p-value
AR (1)	−0.8638	0.1015	−8.5103	1.1×10^{-16}
MA (1)	0.9122	0.1077	8.4698	1.5×10^{-16}
MA (2)	−0.0261	0.0520	−5.0192	3.7×10^{-7}

根据表 2，参数检验在显著性水平($\alpha = 0.05$)下，参数的 T 统计量 P 值都小于 α，可以拒绝原假设，说明参数是显著的。

(2) 对 ARIMA (1，1，2)模型显著性检验。

对残差序列进行 Ljung 白噪声检验，得出六阶延迟下 LB 统计量的 P 值为 0.7993，十二阶延迟下 LB 统计量的 P 值为 0.8879。在显著性水平($\alpha = 0.05$)，P 值都大于 α，所以不能拒绝原假设，说明残差序列是属于白噪声序列，不存在相关的关系，拟合的 ARIMA (1，1，2)模型有效。可以得到以下最终数学模型：

$$x_t = -0.8638x_{t-1} + \varepsilon_t - 0.9122\varepsilon_{t-1} + 0.2610\varepsilon_{t-2} \tag{3-2}$$

3.2.5. 模型预测

将以上通过检验的 ARIMA (1，1，2)模型进行对 35 个交易日沪深 300 指数日收盘价预测。

对 2021 年的 1 月到 2 月交易日做整体预测见表 3。

Table 3. Comparison of some actual values with predicted values
表 3. 部分实际值与预测值比较

时间	真实值	预测值	误差	误差比
2021/1/4	5267.71	5224.65	43.06	0.008176
2021/1/5	5368.50	5219.13	149.37	0.027824
2021/1/6	5417.66	5232.52	185.14	0.034175
2021/1/7	5513.65	5229.58	284.07	0.051522

续表

2021/1/8	5495.43	5240.74	254.69	0.046346
2021/1/11	5441.15	5239.72	201.43	0.037021
2021/1/12	5596.35	5249.22	347.13	0.062028

发现整体预测效果不佳，通过表 3 观察前七天结果，发现预测精度会逐渐变弱，第一天的预测误差低于 1%，预测精度很高，效果是最佳的。

由于以上预测效果不佳，此时考虑使用同一个模型循环预测。也就是说，把训练集数据每预测一次就剔除训练集前面第一个数据，用测试集的第一个数据接在训练集最后，这样保证训练集样本不会变，保持模型参数不变，这样不断迭代 35 次。虽然不能保证每次拟合模型是最佳，但是短期预测影响不大，预测精度还是很高。

35 个交易日的预测值与真实值比较，如图 4：

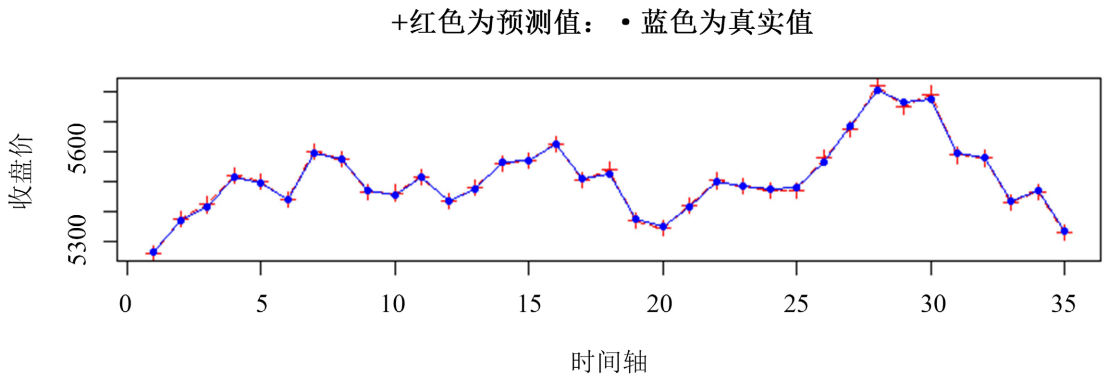

Figure 4. Comparison between predicted values and actual values
图 4. 预测值与真实值比较

3.2.6. 模型效果评价

根据图 4，直观地可以看出循环预测的效果较佳，精度也很高。下面客观的使用平均绝对误差 MAE，平均绝对百分比误差 MAPE 和可绝系数 R^2，判断模型预测的效果。

平均绝对误差：

$$MAE = \frac{1}{n}\left|X_i - \hat{X}\right| \tag{3-3}$$

平均绝对百分比误差：

$$MAPE = \frac{\frac{1}{n}\left|X_i - \hat{X}\right|}{X_i} \times 100\% \tag{3-4}$$

可绝系数：

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3-5}$$

根据以上的三个公式，将 2021 年的 1、2 月收盘价的预测值和真实值在 EXCEL 的简单操作下可以得出下表 4：

Table 4. Model prediction evaluation metric value
表 4. 模型预测评价指标值

MAE	MAPE (%)	R^2
6.0230	0．1085%	0.9959

将 ARIMA (1,1,2)模型进行收盘价预测，平均绝对误差只有 6.0230，平均绝对百分比误差 0．1085%，所有的绝对百分比误差都小于 0.2%，同时可绝系数达到 99%以上，说明该模型的预测效果是非常稳定，预测精度非常高。

4. 总结

本文运用了成熟的时间序列 ARIMA 模型研究沪深 300 指数，对 2021 年 1、2 月开盘日沪深 300 指数收盘价进行预测，预测的准确率经过检验后，预测结果非常好。该文章预测效果相比大多数研究者使用时间序列模型预测股市效果都要好，同时相比大部分研究单一模型预测沪深 300 指数收盘价效果都好。表明使用 ARIMA 模型循环预测，在短期预测效果比较好。

但是，股市是具有不稳定性，股市波动受多种因素影响，同时股市一般不会出现线性趋势，所以单一使用 ARIMA 模型预测，有一定的风险。

基金项目

荆楚理工学院校级科研项目(项目编号：ZKQN2502)。

参考文献

[1] 冯宇旭, 李裕梅. 基于 LSTM 神经网络的沪深 300 指数预测模型研究[J]. 数学的实践与认识, 2019, 49(7): 308-315.

[2] 孙武彪. 基于 BP 神经网络的沪深 300 指数预测分析[D]: [硕士学位论文]. 武汉: 中南财经政法大学, 2019.

[3] 李冰. 沪深 300 股指预测——基于 ARIMA 模型和人工神经网络模型相结合的方法[D]: [硕士学位论文]. 广州: 暨南大学, 2016.

[4] 朱文秀. 基于时间序列分析的沪深 300 指数收盘价预测分析[D]: [硕士学位论文]. 济南: 山东大学 2020.

[5] 徐鑫. 基于 Copula-ARIMA-GJR-GARCH 模型的股票指数相关性分析[D]: [硕士学位论文]. 北京: 清华大学 2015.

[6] 谢太峰, 王硕, 苏磊. 我国股指期货加大了现货市场的波动性吗?——基于 ARMA-GARCH 模型的实证检验[J]. 金融理论与实践, 2017(8): 13-18.

[7] 胡静. 对我国股市预测中 ARIMA-NN 混合模型与 GARCH 族模型的比较研究[D]: [硕士学位论文]. 天津: 天津财经大学 2013.

[8] 张颖超, 孙英隽. 基于 ARIMA 模型的上证指数分析与预测的实证研究[J]. 经济研究导刊, 2019(11): 131-135.

[9] 万建强, 文洲. ARIMA 模型与 ARCH 模型在预测方面的应用比较[J]. 数理统计与管理, 2001(6): 1-4.

[10] 王燕. 时间序列分析——基于 R [M]. 北京: 中国人民大学出版社, 2015.