

# 基于时间序列模型对中国银行股价的预测分析

王浩

贵州大学经济学院, 贵州 贵阳

收稿日期: 2024年3月1日; 录用日期: 2024年4月23日; 发布日期: 2024年5月31日

## 摘要

股价是衡量一家公司在股票市场上的价值和表现的关键指标之一, 也是投资者做出买卖决策的基本依据之一, 可靠的投资建议能有效提升投资者在股票市场的收益。本文尝试运用时间序列模型ARIMA模型, 基于中国银行2018年12月17日至2023年7月31日日交易收盘价数据, 对其股价进行分析与预测。结果显示, 时间序列模型ARIMA可以很好的预测短期内的股价变化, 但股市的影响因素众多, 后续更需要更深入的研究, 发现更适合的模型进行长期预测。

## 关键词

股价, 时间序列, ARIMA模型

# Prediction and Analysis of Bank of China Stock Price Based on Time Series Models

Hao Wang

School of Economics, Guizhou University, Guiyang Guizhou

Received: Mar. 1<sup>st</sup>, 2024; accepted: Apr. 23<sup>rd</sup>, 2024; published: May 31<sup>st</sup>, 2024

## Abstract

Stock price is one of the key indicators used to measure a company's value and performance in the stock market, serving as a fundamental basis for investors to make buying and selling decisions. Reliable investment recommendations can effectively enhance investors' returns in the stock market. This study attempts to employ the time series model ARIMA to analyze and forecast the stock price of China Bank using daily closing price data from December 17, 2018, to July 31, 2023. The results indicate that the ARIMA time series model can effectively predict short-term fluctuations in stock prices. However, considering the multitude of factors influencing the stock market, further in-depth research is essential to identify more suitable models for long-term forecasting.

文章引用: 王浩. 基于时间序列模型对中国银行股价的预测分析[J]. 电子商务评论, 2024, 13(2): 3188-3195.

DOI: 10.12677/ecl.2024.132392

purposes.

## Keywords

Stock Price, Time Series, ARIMA Model

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

股票市场的波动与股价的变化一直以来都是引起广泛关注与研究的课题。股市的不断波动既是市场经济体系中的常态，又是投资者和决策者面临的挑战之一。在全球化、数字化的时代，股市的波动性更加复杂而敏感，受到多种因素的交织影响，包括宏观经济环境、国际政治局势、公司业绩、科技创新等。股价的变动是股票市场最直观、最引人注目的表现之一。股价的涨跌不仅关系到投资者的财富，更直接影响到企业的估值、融资成本以及整个金融体系的稳定性。投资者们时刻关切着市场的脉搏，期望捕捉到股价波动的规律。股票价格历史变动依靠股票价格指数和平均数来记录，也是重要的工具，备受人们关注。一些投资者认为股票价格的变化趋势具有一定可预测性，通过分析历史数据可以提供指引，帮助预测未来股票价格的走势，进而给投资者提供股价波动信息及买卖的合适时机。在股价预测中，常见的模型有时间序列模型 MA，主要利用历史数据的移动平均值来预测未来的股价走势；ARMA 模型结合自回归和移动平均的方法，主要适用于平稳时间序列的预测。后续结合统计学，自回归移动平均模型 (ARIMA) 结合了自回归、积分和移动平均的模型，适用于非平稳时间序列的预测，进一步丰富了时间序列模型。这些方法各具特点，适用于不同的市场情境和数据特征。例如，移动平均法通过对一系列连续时间段的平均值进行计算，平滑了短期内价格波动，突显了长期趋势；而 ARIMA 模型则结合了自回归和移动平均的概念，适用于对时间序列中趋势和季节性进行分析的场景。在实际经济应用中，时间序列分析方法常通过历史数据分析来生成模拟结果，或者通过历史数据预测未来价格变动，从而在特定时间点观测得到的价格数值。ARIMA 模型作为主要的时间序列模型之一，具备优越的短期股票价格变动预测能力，为股票投资者和管理层创造更大的利润提供了有效手段。随着信息时代的到来，python 这一统计分析工具逐渐展露其优势，因此本文选择 python 软件完成 ARIMA 模型的建立及分析，这也拓展了目前对该研究工具的丰富性。

基于上述分析，本文选用中国银行(601988)股票在 2018 年 12 月 17 日至 2023 年 7 月 31 日日交易收盘价数据，利用 python 软件进行平稳性检验，接着对模型进行识别与定阶，选择最佳时间序列模型(本文选用 ARIMA 模型)并对未来交易日股票收盘价进行分析预测，判断其未来一段时间的走势，旨在能为投资者合理买卖股票提供方向和参考价值。

## 2. 文献综述与模型理论

### 2.1. 文献综述

股票市场预测自 19 世纪问世以来一直是学者们关注的焦点。早期，学者们提出了多种分析方法，如道氏分析法、K 线图分析法、趋势分析法、四度空间法等等。在这些方法的基础上，预测股价的研究不断发展。在 1933 年，Alfred (1944)在《Stock Market Forecasting》一书中首次提出了股市导向预测的概念，

这可以看作是对股价预测的初步尝试。后来，一些研究者尝试了不同的方法。比如，王培勋[1]利用最小二乘法拟合的成交量回归方程来预测股市，但是这种方法的预测结果并不尽如人意。随后，刘红梅[2]运用 ARMA (2, 1)模型来预测分析鞍钢股份的股价走势，这种尝试使得预测效果有所提升。然而，虽然取得了一定进展，但预测的精度仍有待提高。

股价受多种因素影响，使得股价序列内部规律复杂，难以用传统方法简单建模。因此，传统方法在股价预测方面通常难以取得理想效果。人工智能的崛起，时间序列分析和神经网络走进人们视野。GenCay [3]在预测道琼斯工业平均指数时，使用了 BP 神经网络，将移动平均股价作为输入变量。结果表明，其预测效果明显优于线性模型。Merh [4]等尝试构建了一个综合了三层前馈反向传播神经网络(ANN)和自回归综合移动平均线(ARIMA)的混合模型，用于对印度股市未来指数值和趋势进行预测。他们通过比较混合模型的仿真结果与基于 ANN 的模型以及基于 ARIMA 的模型的结果，来评估混合模型的性能和收敛性。这种混合模型可以为股市预测提供更为综合和准确的信息。高天[5]采用了最优小波包变换技术对股票价格进行分解，将其拆分成不同的成分，并利用 ARIMA 模型预测短期趋势。陈小玲[6]研究百度和阿里巴巴股票收盘价序列，分别将 ARIMA 模型和 BP 神经网络技术融合到股价预测的研究中，比较和验证这两种模型在股价预测方面的有效性和准确性，通过对比分析来评估它们在预测精度方面的表现。杨宇焜[7]利用 ARIMA 模型对大众公用开盘价进行预测，验证该模型预测误差小，具有一定参考价值。王源和李俊刚[8]对上证指数历史数据进行分析，比较了移动平均法、自回归综合移动平均和 LSTM 三种算法在股票走势预测中的精度，发现 LSTM 算法更准确预测短期股价。熊政和车文刚[9]采用时间序列组合模型 ARIMA-GARCH-M 对短期股价进行预测，取得了良好的效果，具有可行性。翁紫霞[10]对建设银行收盘价数据使用时间序列模型 ARIMA 进行预测和分析，结果显示短期预测效果较好，但长期预测误差较大，提示需要探索更为准确的股价预测模型。

综上所述，自那时以来，股票市场预测一直是一个极具挑战性的领域，研究者在不断探索和改进各种模型和方法，以提高对股价走势的准确性和预测精度。

## 2.2. 模型理论

### 2.2.1. ARMA 模型介绍

ARMA 模型称为自回归移动平均模型，该模型由自回归模型(AR 模型)和移动平均模型(MA 模型)组成的，用于进行平稳时间序列分析和预测。首先，AR 模型是一种利用时间序列自身的过去观察值来预测未来值的模型。它基于时间序列内部的自相关关系，使用过去时刻的观察值与现在的干扰项(误差项)的线性组合来进行预测。AR 模型的核心思想是当前值与过去值之间存在一定的线性关系。AR ( $p$ )模型形式表达为：

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t$$

式中， $y_t$ 为平稳时间序列， $\varepsilon_t$ 为误差， $\Phi$  ( $i = 1, 2, \dots, p$ )表示 AR 模型的待定系数， $p$ 表示 AR 模型的阶数。

MA 模型是一种利用时间序列的预测误差项的线性组合来进行预测的模型。它假设当前观察值与先前的干扰项之间存在相关性，通过对这些干扰项的加权平均来进行预测。MA( $q$ )模型形式表达为：

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

式中， $y_t$ 为平稳时间序列， $\varepsilon_t$ 为误差， $\theta$  ( $j = 1, 2, \dots, q$ )表示 MA 模型的待定系数， $q$ 表示 MA 模型的阶数。

ARMA 模型结合了这两种模型的优势，既考虑了过去观察值的影响(AR 部分)，又考虑了过去误差的影响(MA 部分)。这种结合使得 ARMA 模型能够更好地描述许多实际时间序列数据的特征，并用于预测

未来的值。ARMA ( $p, q$ )模型形式表达为:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

### 2.2.2. ARIMA 模型

ARIMA 模型是一种经典的时间序列分析方法,由博克思(Box)和詹金斯(Jenkins)在 20 世纪 70 年代初提出。这个方法用于预测时间序列数据的未来趋势,结合了自回归(AR)和移动平均(MA)的成分,同时考虑了数据的差分(Integrated),因此得名 ARIMA。ARIMA 是 ARMA 模型的衍生,二者主要不同点是 ARMA 模型仅适用于该时间序列是平稳的,而 ARIMA 模型既适用于平稳也适用于非平稳时间序列。在实际应用中,由于许多时间序列是非平稳的,ARIMA 模型的应用更为广泛。ARIMA 模型的参数通常由三个部分表示:AR (自回归)、I (差分阶数)、MA (移动平均)。通过分析自相关函数(ACF)和偏自相关函数(PACF),可以确定这些参数的最佳值。

ARIMA ( $p, d, q$ )模型中,AR 代表自回归, $p$  表示自回归的阶数;MA 代表移动平均, $q$  表示移动平均项数;而  $d$  则表示时间序列进行平稳化处理时的差分次数(阶数)。因此,ARIMA ( $p, d, q$ )模型形式可以简洁地表示为:

$$\Delta d y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

式中, $\Delta d y_t$  表示不平稳序列  $y_t$  经过  $d$  次差分转换形成的平稳序列, $\varepsilon_t$  为误差, $\Phi$  ( $i = 1, 2, \dots, p$ )和  $\theta$  ( $j = 1, 2, \dots, q$ )为模型的待定系数, $p$  和  $q$  为模型的阶数。

### 2.2.3. ARIMA 模型建模

ARIMA 模型建模主要分为以下三个步骤:

第一:预处理数据,将不平稳数据差分处理,使其趋于平稳化,最后进行平稳性检验。绘制中国银行历史收盘价数据的时间序列图,将其进行一阶差分处理并趋于平稳化。

第二:识别模型和检验模型。绘制自相关系数和偏自相关系数函数分析图,确定模型参数  $p$  (自回归阶数)、 $q$  (移动平均阶数)、 $d$  (平稳化过程的差分次数),选择合适参数拟合模型。

第三:预测。建立模型后对中国银行未来一周为间隔的五次股价进行预测,将预测值与实际收盘价比较,算出相对误差,进而分析短期和长期预测的效果。也可反向验证选择该模型进行预测是否合适。

## 3. 实证分析

### 3.1. 数据选取及来源

股票通常参考开盘价、收盘价、最高价、最低价四个指标中进行预测,而股市中的股价都是时变的,某一时间点上的不同供求关系就通过最高价和最低价显示,因此这两者不适合进行股票的总体预测。开盘价就是前一日的收盘价,代表着前一天的供求关系,而收盘价代表整个交易日结束的价格,也是股票投资者对整日价格动向的最终判断,因此本文选择收盘价进行股票的预测变量。

中国银行作为中国唯一持续经营超过百年的银行,展现了卓越的韧性和持续发展的能力。作为国际化和多元化程度最高的银行之一,其在 A 股市场上的表现也一直相对稳健。该银行在金融领域的强大地位使其具有很高的代表性。因此本文选择中国银行(601988)股票在 2018 年 12 月 17 日至 2023 年 7 月 31 日日交易收盘价数据,共 1121 组观测数据,数据来自 CSMAR 数据库。

### 3.2. ARIMA 模型建立

#### 3.2.1. 序列平稳化处理和平稳性检验

本文用 python 软件绘制中国银行(601988)股票在 2018 年 12 月 17 日至 2023 年 7 月 31 日的时间序

列图, 图 1 为中国银行原始收盘价时间序列图, 其显示中国银行季节性分布趋势不显著, 2019 年下半年到 2023 年初几乎处于下降趋势, 2023 年以后大幅度上升, 波动幅度较大, 属于不平稳时间序列。因此对原始数据进行差分处理。时序图(图 2)显示数值大部分落在 0 上下, 存在小幅度波动, 基本判定为该时间序列趋于平稳化。



Figure 1. Original closing price time series chart

图 1. 原始收盘价时间序列图

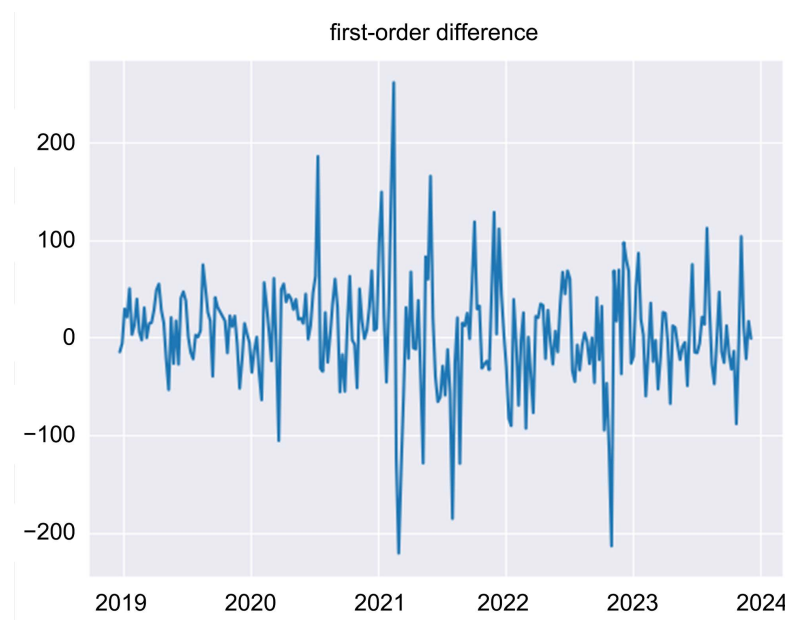


Figure 2. Time series chart of closing price after 1st order differential processing

图 2. 经 1 阶差分处理后收盘价时间序列图

### 3.2.2. ARIMA 模型识别与定阶

经过一阶差分后得到平稳时间序列，本文拟选用 ARIMA ( $p, d, q$ )模型进行模型识别并进行定阶，定阶需要将  $p, d, q$  三个参数确定下来。首先这是对原始序列进行一阶差分获得的平稳序列，可以确定  $d=1$ 。其次，根据自相关图(ACF)和偏自相关图(PACF)来主观确定  $p, q$  值(如图 3、图 4 所示)。1 阶差分自相关系数 ACF 图和偏自相关系数 PACF 图均显示为拖尾，证实该平稳时间序列适用于 ARIMA 模型。再次分析图 3、图 4，可以得到系数  $p=2$  (自回归阶数)， $q=3$  (移动平均阶数)，从而本文确定模型参数为 ARIMA (2, 1, 3)，本文基于此建立方程。

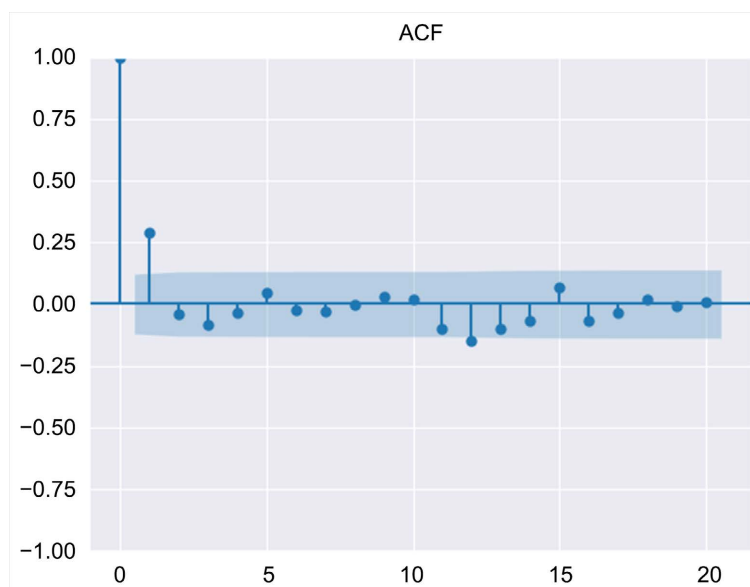


Figure 3. ACF diagram  
图 3. ACF 图

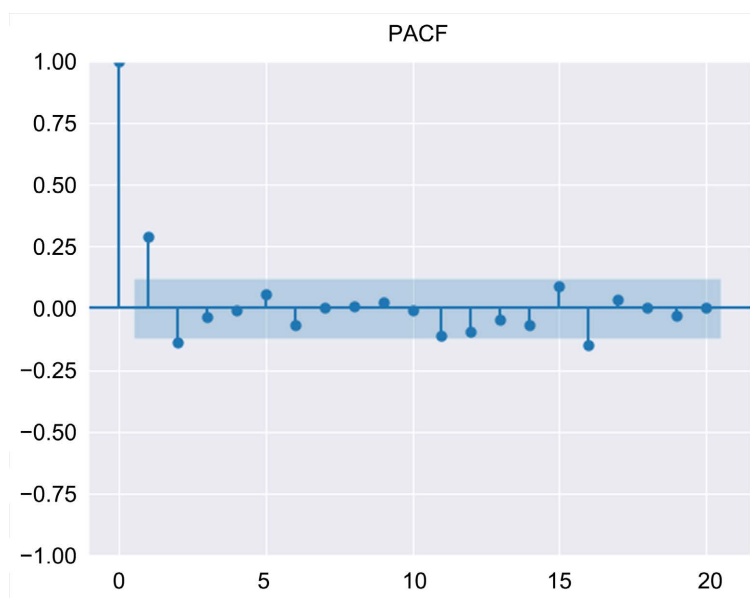


Figure 4. PACF diagram  
图 4. PACF 图

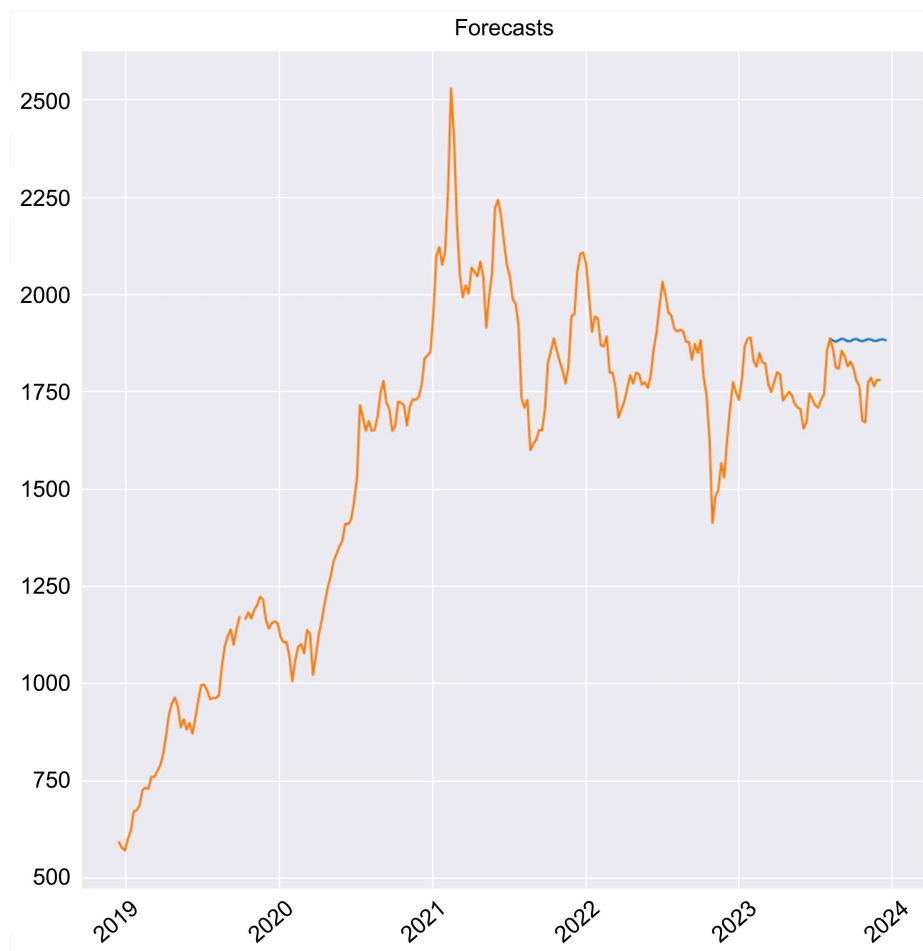
### 3.2.3. ARIMA 模型预测

利用前面已建立的 ARIMA (2, 1, 3)模型预测中国银行的收盘价，由于股市的波动是时变且较大的，因此本文仅进行短期的预测以求较为准确的数值。以下是未来间隔一周的 5 个交易日收盘价预测结果，如表 1 所示。

**Table 1.** Forecast results of closing price for the next 5 trading days with a one week interval

**表 1.** 间隔一周未来 5 个交易日收盘价预测结果

交易日	预测值	实际收盘价	误差(%)
2023.8.7	3.886	3.86	0.67
2023.8.14	3.876	3.72	4.19
2023.8.21	3.843	3.73	3.03
2023.8.28	3.861	3.84	0.55
2023.9.4	3.882	3.74	3.79



**Figure 5.** Stock forecast results

**图 5.** 股票预测结果

根据表 1 和图 5 的结果所示，运用该模型进行中国银行短期的股价预测值与实际收盘价较为接近，其误差都未超过 5%，尤其是短期内，说明该模型模拟效果较好且短期预测精度较高。预测的股价能较好

模拟出中国银行股价短期下降趋势。然而从图 5 的后半段可以看出,如果预测长期股价,ARIMA (2, 1, 3) 模型的预测效果与实际收盘价的走势有较大差别,效果并不理想。

#### 4. 结论与建议

本文运用 python 软件对中国银行股票收盘价这一时间序列进行 ARIMA 模型构建与实证分析,最终得出:该模型短期拟合结果较为准确,解决了非平稳时间序列问题,但长期预测效果并不理想。时间序列分析为投资者提供了一种从历史数据中提炼规律的手段,但决策的最终结果仍然需要综合考虑市场的多变因素。

例如大盘走势对股价的影响是非常巨大的。同时,中国的股票市场一直处于动态且不断变化的状态,受多方面因素的综合影响,包括新冠疫情、宏观经济调控政策以及相关法规的颁布等。投资者在使用时间序列分析进行股价预测时,应当清醒认识到模型的局限性,以及市场变化的不确定性。合理的投资策略应该是基于对市场宏观环境、行业发展、公司基本面等多方面因素的全面分析,而非仅仅依赖于模型的结果。综合分析这些因素,投资者可以借助时间序列模型评估和判断特定股票在短期内的盈亏变化。这有助于投资者理性看待股票市场中可能存在的负面影响,从而制定更为精准的短期投资计划,使其在投资决策中更具科学性。然而,对于长期投资而言,建议投资者深入研究并寻找更准确的股价预测模型,以更好地应对市场的复杂性和不确定性。

#### 参考文献

- [1] 王培勋. 非线性回归的弹性分析在股票投资与行情预测中的应用[J]. 系统工程理论与实践, 2000(8): 100-104.
- [2] 刘红梅. ARIMA 模型在股票价格预测中的应用[J]. 广西轻工业, 2008(6): 92-93.
- [3] GenGay, R. (1996) Non-Linger Prediction of Serurns with Moving Average Rules. *Journal of Focceasting*, **15**, 165-174. [https://doi.org/10.1002/\(SICI\)1099-131X\(199604\)15:3%3C165::AID-FOR617%3E3.0.CO;2-V](https://doi.org/10.1002/(SICI)1099-131X(199604)15:3%3C165::AID-FOR617%3E3.0.CO;2-V)
- [4] Merh, N., Saxena, V.P. and Pardasani, K.R. (2010) A Comparison between Hybrid Approches of ANN and ARIMA for Indian Stock Trend Forecasting. *Business Intelligence Journal*, **3**, No. 2.
- [5] 高天. 基于最优小波包变换、ARIMA 与 SVR 的股票价格预测研究[J]. 贵州财经大学学报, 2015(6): 57-69.
- [6] 陈小玲. 基于 ARIMA 模型与神经网络模型的股价预测[J]. 经济数学, 2017, 34(4): 30-34.
- [7] 杨宇源, 张梅. 基于 ARIMA 模型的股票价格实证分析[J]. 科技资讯, 2021, 19(29): 121-123+127.
- [8] 王源, 李俊刚. 预测股票市场走势的模型评估[J]. 中国集体经济, 2022(27): 116-120.
- [9] 熊政, 车文刚. ARIMA-GARCH-M 模型在短期股票预测中的应用[J]. 陕西理工大学学报: 自然科学版, 2022, 38(4): 69-74.
- [10] 翁紫霞. 基于 ARIMA 模型的股价分析与预测——以建设银行为例[J]. 现代信息科技, 2023, 7(14): 137-141.