

ARIMA模型在股价预测上的应用

——以中国银行为例

陈德慧

上海工程技术大学管理学院, 上海

收稿日期: 2023年4月13日; 录用日期: 2023年6月27日; 发布日期: 2023年7月5日

摘要

随着我国资本市场的扩大开放, 国外资本进一步流入我国股市, 中国股市迎来新的挑战与机遇。金融市场股票的价格预测问题再次成为公众关注的热点。基于此, 选取金融市场股票的历史收盘价数据, 以R为实现工具, 通过建立ARIMA模型来进行分析与预测, 为股市投资者提供帮助。

关键词

ARIMA模型, 股价, 短期预测

Application of ARIMA Model in Stock Price Prediction

—Taking Bank of China as an Example

Dehui Chen

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Apr. 13th, 2023; accepted: Jun. 27th, 2023; published: Jul. 5th, 2023

Abstract

With the expansion and opening up of China's capital market, foreign capital further flows into China's stock market, and China's stock market faces new challenges and opportunities. The price prediction of stocks in financial markets has once again become a hot spot of public concern. Based on this, this paper selects the historical closing price data of stocks in the financial market, and uses R as the realization tool to analyze and predict by establishing ARIMA model, so as to provide help for stock market investors.

Keywords

ARIMA Model, Stock Price, Short-Term Forecast

Copyright © 2023 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

股票市场在金融领域中一直有着举足轻重的地位。股票价格的波动能够在一定程度上反映一个国家经济周期的变化。近年来随着经济的发展和疫情的冲击，股票的价格再次引发了大家广泛的关注。韩金磊等采用长短期记忆网络(LSTM)和灰色模型对股价进行预测研究[1]；荆思寒等基于 A 股上市公司股票交易数据建立了股票风险传染模型[2]；左川等记录了深度学习在股票投资领域中的应用[3]。

股票是经济的晴雨表，无论是对于国家还是对于投资者来说股价预测都具有重要的意义[4] [5]。股票价格往往对时间因素非常敏感，因此本文建立 ARIMA 时间序列模型对短期股票价格进行分析与预测。

2. 模型的建立

2.1. ARIMA 模型的定阶方法

ARIMA 模型的定阶方法主要是通过观察自相关图和偏自相关图，根据其系数的相关特性来判断阶数。若 AC 是拖尾的，PAC 是 p 阶截尾的，则选择 AR(p)模型；若 PAC 是拖尾的，AC 是 q 阶截尾的，则选择 MA(q)模型；若 AC 和 PAC 都表现为拖尾性，则选择 ARMA(p,q)模型，然后通过模型的参数检验，进一步判断 p、q 的阶数。

由于样本具有随机性，相关系数并不会表现出完美截尾的情况，通常会呈现出小值振荡的情况，这就给模型的定阶带来了一定的困难。根据以往的经验，如果样本自相关系数或偏自相关系数在初始的 d 阶明显大于两倍标准差范围，其后几乎 95%的自相关系数都在两倍标准差范围内，并且由非零自相关系数衰减为小值波动的过程非常突然，这时可以视为自相关系数或偏自相关系数 d 阶截尾。

2.2. ARIMA 模型建模步骤

构建 ARIMA 模型的大致步骤有以下几个：

(一) 序列的平稳性。首先，需要观察原序列是不是平稳的[6]。当原序列不是平稳的情况时，可以选择一些合理的方法让原序列变成平稳的，如差分转化。

(二) 模型的定阶。通过软件能够计算得出序列特征的自相关系数与偏自相关系数等。ARIMA 模型的滞后阶数 p 和 q 再利用其他方法确定。

(三) 模型的构建。构建该模型的残差序列是否是正态分布的也值得关注。残差序列的自相关图以及偏自相关图可以观察得出其结论，或运用检验序列相关的方法来测试残差序列的随机性。

(四) 模型的预测。经过了上面的步骤后就大致拟合出了模型，并可以在未来的短期时间里进行股价的预测[7]。

3. 关于股票价格的预测

选取中国银行(601988) 2022 年 4 月 10 日至 2023 年 4 月 10 日的股票日开盘价数据。基于 ARIMA 模

型的建模理论，建立阶数 $p = 0, d = 1, q = 1$ 的 ARIMA(0,1,1)模型对中国银行未来三天的股票日开盘价进行预测[8]。

3.1. 平稳性检验

3.1.1. 时序图检验

从图 1 的折线图大概可以推测出该序列数据可能并不是平稳的。因为股票的开盘价数据为原始数据，该数据在短期内波动是比较大的，在长期的时间序列中也没有遵循显著的规律[9] [10]。

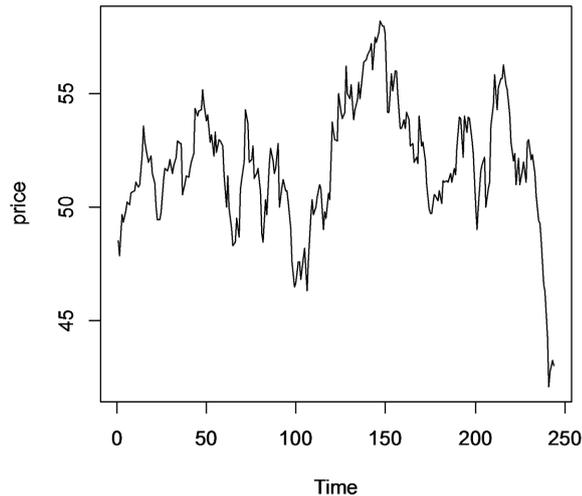


Figure 1. Timing diagram of opening price of China Bank
图 1. 中国银行开盘价时序图

3.1.2. 单位根(ADF)检验

为了更加精准得确认该序列是不是平稳的，对序列做 ADF 单位根检验。在显著性水平下，该序列的 P 值大于 0.05，因此该时间序列的数据不是平稳的。

对原始数据的时间序列进行一阶差分处理后，将一阶差分后的数据再次进行 ADF 检验。序列的 p 值小于显著性水平 0.05，因此推断出该序列属于平稳的时间序列。对于一阶差分后的数据再进行白噪声检验，该序列为非白噪声系列。综上，对以上结果分析该序列为平稳非白噪声序列。

3.2. 序列的自相关图和偏自相关图并识别模型

在对时间序列数据进行平稳性处理之后，需要利用 ACF、PACF 图，如图 2、图 3 所示来识别模型形式，并确定滞后阶数 p, q 的值。

观察样本自相关图和偏自相关图，如图 2、图 3 所示。由图 2 可知自相关系数除了滞后一阶的数值较大，其他数值均在两倍标准差范围内做小值随机波动，并且相关系数衰减的过程非常突然，所以可以判断自相关系数为一阶截尾。由图 3 可知，偏自相关系数缓慢衰减且过程相当连续，呈现拖尾性质。

综上，决定建立 ARIMA (0,1,1)模型来预测该序列，即模型的阶数 $p = 0, d = 1, q = 1$ 。但此推测仍然具有一些主观性，所以需要进一步检验来验证[11]。

3.3. 模型的拟合性检验

完成上述参数估计步骤后，最关键的步骤就是要对该拟合模型的残差序列进行正态性检验[12]。我们

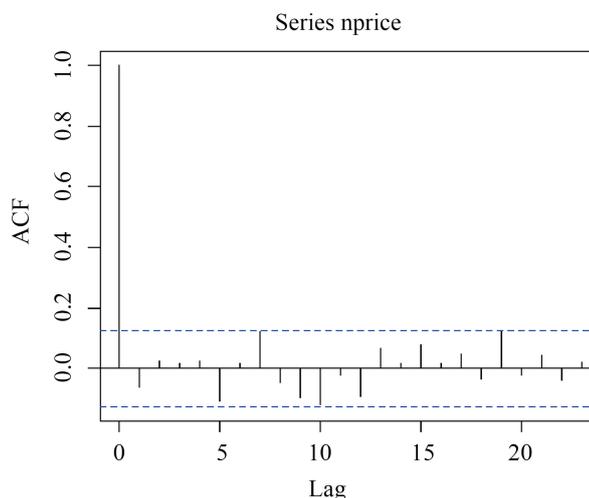


Figure 2. ACF diagram of first-order differential data

图 2. 一阶差分后数据的 ACF 图

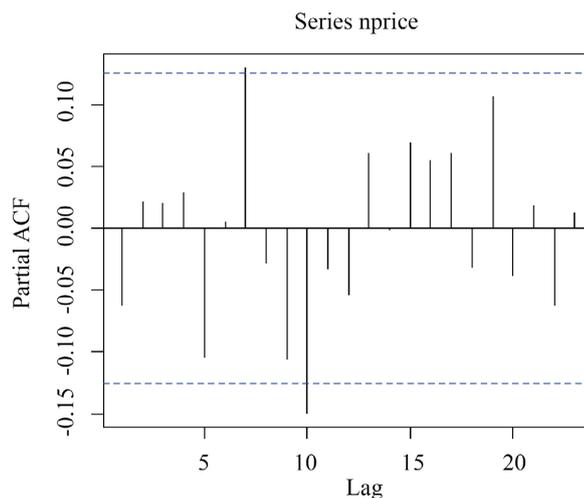


Figure 3. PACF diagram of first-order differential data

图 3. 一阶差分后数据的 PACF 图

在做序列拟合的时候， Y 应该是 X 的函数。但是现实并没有那么友善，往往会出现一些奇奇怪怪的事让 X 与 Y 不符合这个函数。比如对身高的影响，一对双胞胎，从小一块长大，但是他们身高不可能完全一样。所以把这种只有上帝才知道怎么回事的变化叫做随机误差项，它本身就是随机的、不可预测的。

我们所做的模型中的残差就是对这个随机误差的估计，残差如果不是随机的，那说明残差中的一部分还是与预测变量有关系，需要再继续拟合。如果它是随机的就可以认为它是对随机误差比较好的拟合，就可以认为你的可确定部分 X 提取了所有的可预测部分，证明你的模型是把所有的 X 提取全了。我们对文中所建模型的残差序列进行正态性检验，发现其残差符合正态分布。表明所建模型拟合程度较好，可以较为准确地预测未知数据。

3.4. 对开盘价进行预测分析

下面利用所构建的 ARIMA 模型对中国银行股票的日开盘价进行预测。由于 ARIMA 模型短期内预测

的结果较为精确，因此本文只是对股票价格做出短期限内的预测，即 2023 年 4 月 11 日至 2023 年 4 月 13 日的开盘价。如表 1 所示：

Table 1. Opening price and forecast price results

表 1. 开盘价与预测价结果

日期	开盘价	开盘价预测	误差比
2023.4.11	3.43	3.15	0.0816
2023.4.12	3.46	3.18	0.0809
2023.4.13	3.49	3.24	0.0716

通过表 1 可以观察到，利用这个模型预测的中国银行股票 2023 年 4 月 11 日至 2023 年 4 月 13 日的开盘价分别是 3.43、3.46、3.49，误差比分别为 0.0816、0.0809、0.0716。我们发现实际值与预测值之间的误差比是比较小的，根据上述数据进一步证实了此次构建的模型是准确的。

4. 结语与建议

本文对中国银行股票开盘价这一时间序列进行 ADF 单位根检验，绘制该序列的自相关图和偏自相关图。建立了阶数 $p=0$ ， $d=1$ ， $q=1$ 的 ARIMA(0,1,1)模型对股票开盘价进行预测。由于 ARIMA 模型不能精确拟合股票价格的长期趋势，只适用于短期预测。于是我们运用该模型预测中国银行未来三天的股票开盘价[13]。模型预测中国银行股票 2023 年 4 月 11 日至 2023 年 4 月 13 日的开盘价分别是 3.43、3.46、3.49。真实值与预测值的误差比分别为 0.0816、0.0809、0.0716。误差比较小，因此所构建的股票开盘价预测模型合理。

文末，给资本市场一些建议。

4.1. 健全信息披露机制

我国市场存在十分典型的羊群效应。捕风捉影投资方式大量存在，即股民们根据从各个渠道收集来的内幕消息买进卖出股票。若程度过大就将造成市场极大的波动，股民们对市场失去信心势必会影响家庭资本结构的变化，同时内需也得不到刺激。健全信息披露机制可以防止股市环境恶化，促进投资者进行理性投资。

4.2. 加强证券市场化建设

在供给侧改革背景下，政府对证券市场的干预不能太大，但减少投机行为的产生也是在当下的实际情况中所必要的。充分发挥市场作用、加强市场化建设是相关部门需要进一步完善的。

4.3. 对投资者的理性教育

股市中的投资者以散户居多，这导致市场中跟风投资的现象比较明显，所以对投资者进行理性教育是必不可少的。投资者在买卖证券时应该基于其实际情况进行选择。

参考文献

- [1] 韩金磊, 熊萍萍, 孙继红. 基于 LSTM 和灰色模型的股价时间序列预测研究[J/OL]. 南京信息工程大学学报(自然科学版):1-22. <https://kns.cnki.net/kcms/detail/32.1801.N.20230105.1635.003.html>, 2023-04-20.
- [2] 荆思寒, 王振山, 隋聪, 马天一, 任昭诺. 股票间的风险传染——基于对股价崩盘风险的预测[J]. 系统工程理论

与实践, 2022, 42(11): 3090-3104.

- [3] 左川, 王宇, 李振. 深度学习在股票投资中的应用[J]. 郑州大学学报(哲学社会科学版), 2022, 55(2): 56-62.
- [4] 周章元, 何小灵. 基于优化 LSTM 模型的股价预测方法[J]. 统计与决策, 2023, 39(6): 143-148.
- [5] 王瑶, 冯晓晴, 侯德帅. 企业数字化转型能提高分析师预测准确度吗?——基于信息披露和信息挖掘的双重视角[J/OL]. 中南财经政法大学学报: 1-14. <https://doi.org/10.19639/j.cnki.issn1003-5230.20230130.003>, 2023-04-12.
- [6] 黄宇漩, 杨胜刚, 陈收, 朱琦. 证券分析师关注与投资-股价敏感性[J]. 系统工程理论与实践: 2023, 43(4): 1022-1043..
- [7] 林虹, 张永, 杨兴雨, 黎嘉豪. 考虑组合预测股价的泛证券投资组合选择策略[J/OL]. 管理工程学报: 1-12. <https://doi.org/10.13587/j.cnki.jieem.2023.05.011>, 2023-04-12.
- [8] 王满, 张苗苗. 考虑高维宏观信息的波动率与股票价格预测[J]. 统计与决策, 2022, 38(20): 138-143.
- [9] 牛桂草, 周绩宏, 马红燕, 杜晓东, 王俊芹. 基于 ARMA 模型的河北鸭梨价格预测及预警[J]. 山东农业科学, 2021, 53(11): 151-156.
- [10] 王冠智, 栗晓玲, 张特, 姜田亮, 褚江东. 基于 DWT-WFGM(1,1)-ARMA 组合模型的农业用水量预测[J]. 灌溉排水学报, 2021, 40(11): 106-114.
- [11] 黄卫, 刘升. 基于长期线性趋势的时间序列建模研究[J]. 统计与决策, 2021, 37(2): 30-33.
- [12] 冯盼, 曹显兵. 基于 ARMA 模型的股价分析与预测的实证研究[J]. 数学的实践与认识, 2011, 41(22): 84-90.
- [13] 谢赤, 喻欣欣, 李双飞. 双重上市公司 A+H 股价差时间序列描述与趋势预测[J]. 当代财经, 2008(5): 57-60.