

基于 NYSE 股票价格的时间序列分析研究

罗兰一星

中央财经大学 北京

【摘要】随着如今金融市场的不断发展，股票价格变动成为人们关注的焦点。股票价格不仅反映了公司的经营状况和市场价值，还受到宏观经济、政策调整、等多种因素的影响。因此，如何对股票价格进行准确地分析和预测，对于指导投资决策、维持金融市场稳定等具有重要意义。本文以 1984-2012 长时间跨度的纽约证券交易所美股收盘价格数据为基础，从分层和整体两方面构建时间序列模型对股票价格进行预测和分析。在实证分析当中，通过对比不同模型方法，旨在为提高股票价格预测的准确性和可靠性提供新的思路和方法，从而为投资者与金融机构提供更准确的决策支持，促进金融市场的稳定和发展。

【关键词】时间序列；ARIMA 模型；股票价格分析

【收稿日期】2024 年 5 月 18 日 **【出刊日期】**2024 年 6 月 21 日 **【DOI】**10.12208/j.aam.20241006

Research on time series analysis of NYSE stock prices

Lanyixing Luo

Central University of Finance and Economics, Beijing

【Abstract】 With the continuous development of the financial market, stock price fluctuations have become the focus of people's attention today. Stock prices not only reflect a company's operating conditions and market value, but are also influenced by various factors such as macroeconomic and policy adjustments. Therefore, accurate analysis and prediction of stock prices are of great significance for guiding investment decisions and maintaining financial market stability. Based on the close price data of the New York Stock Exchange from 1984 to 2012, this paper constructs a time series model to forecast and analyze stock prices from two aspects: stratification and whole. In the empirical analysis, by comparing different model methods, it aims to provide new ideas and methods for improving the accuracy and reliability of stock price prediction, so as to provide more accurate decision support for investors and financial institutions, and promote the stability and development of financial market.

【Keywords】 Time series; ARIMA model; Stock price analysis

1 引言

在当今世界，股票市场已成为全球经济的重要组成部分，对各国经济、企业发展和个人财富都有深远的影响，我国股票市场规模也在逐步扩大，蓬勃发展。因此，研究股票具有极其重要的意义。而随着大数据时代的到来，时间序列方法作为研究生态系统的行为、预测未来趋势的强大工具，在股票价格的分析和预测中可以发挥重大作用。

程孟菲、高淑萍^[1]提出多尺度股票价格预测模型，根据分量的震荡频率不同分别选择深度迁移学习与移动平均法进行预测。张艺萧^[7]对 LSTM 神经网络模型进行了详细阐述，并详细描述了其相较自回归模型等其他模型的优势，并对 LSTM 神经网络模型的算法进行了改进，使学习速率 α 变得更加灵活。李奋华、赵润林^[8]分别通过指数平滑预测法和 ARMA 预测法分别预测北京利尔股票数据，研究发现，ARMA 预测法对股票的预测效果远好于指数平滑预测法，说明相较于线性时间序列预测法，非线性时间序列的预测法在实际的股票市场预测中预测效果更好。Zhao 等人^[3]提出基于 RNN 的预测模型，通过对注意力机制的引入来聚

焦关键信息。

杨智勇、叶玉玺、周瑜^[9]针对股票预测模型存在的预测功能单一、时效性问题，以 LSTM 为基础研究得到了双向长短期记忆（BILSTM）神经网络（BILSTM-SA-TCN）股票预测模型。该模型可在有效学习股票数据的同时抓取长时间的依赖信息，具有更稳定的预测效果。高雅婷^[10]在《时间序列趋势方法研究及其在股票市场中的运用》中提出了基于相关熵的回归预测模型，能够充分减少异常数据的影响，预测精度较高。陈登建、杜飞霞、夏换^[11]将 ARIMA 模型与 SVR（滚动残差的支持向量机回归模型）进行组合，实现更准确地预测，但依然存在不能进行长期预测的不足。Mengxia Liang, Shacong Wu, Xiaolong Wang, Qingcai Chen^[12]在《A stock time series forecasting approach incorporating candlestick patterns and sequence similarity》文中基于序列模式与序列相似度，挖掘多元时间序列数据的 k 线模式，并验证出该混合模型相较 ARMA 模型与 LSTM 模型而言具有更高的预测准确度。如 Zhang 等人通过提高注意力的 LSTM 模型与多种来源数据相融合，显著提高了股票预测性能。[2]Pin Lv, Qinjuan Wu, Jia Xu 与 Yating Shu^[13]在《Stock Index Prediction Based on Time Series Decomposition and Hybrid Model》文中提出了一种基于时间序列分解和混合的新股指数预测模型，具体方法是将股票指数利用自适应噪声分析进行的完全集合经验模式分解（CEEMDAN），将股指分解为一系列具有不同特征尺度和趋势项的本征模式函数（IMF），运用 ADF 单位根检验法判断每个 IMF 和趋势项的稳定性。Jakub Horák 与 Tomáš Krulický^[14]在《Comparison of exponential time series alignment and time series alignment using artificial neural networks by example of prediction of future development of stock prices of a specific company》文中，将时间序列的指数排列法与运用人工神经网络法作为预测公司股价工具的时间序列法进行比较，研究发现，传统的时间序列指数排列法的预测结果比运用人工神经网络作为工具的预测方法有更好的预测效果。但在该文章中，作者并未找到导致这一结果的原因，可能的原因包括预测周期较长（62 个工作日）、其选择的公司 Unipetrol 股价较为动荡等，还需进一步研究。

本文的文章结构为：首先在第二部分介绍本文预测股票波动会用到的时间序列模型及方法，在第三部分基于本研究数据结果进行具体分析，最后得出本文结论与研究展望。

2 模型及方法

2.1 AR 模型

AR 模型，即“自回归模型”（Autoregressive Model），是时间序列分析中的一种基本模型之一。

AR 模型的数学表达式如下：

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t$$

其中， X_t 是时间序列在时刻 t 的观测值， $x_{t-1}, X_{t-2}, \dots, X_{t-p}$ 是在 t 之前的 p 个时刻的观测值， $\phi_1, \phi_2, \dots, \phi_p$ 是模型的参数，表示对应时刻的系数， ϵ_t 是误差项。

AR 模型能够捕捉时间序列数据中的自相关结构，且通过拟合历史数据来进行未来值的预测。

在具体实践中，为了确定 AR 模型的阶数，通常会借助自相关函数（ACF）和偏自相关函数（PACF）。通过观察 ACF 图像确定时间序列数据中的整体自相关性，PACF 图像则可以特定时刻与当前时刻之间的直接相关性。ACF 在滞后阶数 p 后截尾，PACF 截尾这表明时间序列的当前观测值与过去 p 个滞后观测值之间存在显著的相关性，而其他滞后期的相关性很小。因此，可以判断出模型为 p 阶 MA 模型。

2.2 MA 模型

MA 模型，即移动平均模型（Moving Average Model），也是一种常用于时间序列分析的统计模型，MA 模型的数学表达式如下：

$$X_t = \mu + \epsilon_t + \vartheta_1 \epsilon_{t-1} + \vartheta_2 \epsilon_{t-2} + \dots + \vartheta_q \epsilon_{t-q}$$

MA 模型的阶数 q 表示模型中考虑的白噪声的数量，称为移动平均阶数。较高的移动平均阶数通常意味着模型考虑了更长的时间范围内的波动。ACF 截尾，PACF 在滞后阶数 q 后截尾时，意味着时间序列的当前观测值与过去 q 个滞后观测值之间存在显著的相关性，而其他滞后期的相关性很小。因此，可以判断

出模型为 q 阶 MA 模型。

2.3 ARIMA 模型

如果 ACF 和 PACF 图中没有明显的截尾或衰减模式,可能表明时间序列不适合简单的 MA 或 AR 模型。这种情况下需要考虑更复杂的模型,例如 ARMA (自回归移动平均) 模型。

形如 ARIMA (p, d, q) 的模型是自回归积分移动平均模型 (Autoregressive Integrated Moving Average Model), 其中 p 表示自回归项的阶数, d 表示差分次数, q 表示移动平均项的阶数。在 ARIMA 模型中, 差分是通过和时间序列进行减法运算来减少数据的趋势和季节性, 使其更加平稳。

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

其中 x_t 是时间序列的当前值, $\phi_1 \dots \phi_p$ 是自回归系数, ϵ_t 是白噪声误差项, $\theta_1 \dots \theta_q$ 是移动平均系数。

本文的建模步骤:

(1) 将 1984-2012 纽约证券交易所美股收盘价格作为研究对象, 画出其时间序列图像, 判断其平稳性。

(2) 对非平稳的数据进行平稳化处理。由于非平稳序列不具有二阶矩平稳的性质, 很难进行拟合, 因此需要进行进一步平稳化处理后确定适当的拟合模型。

(3) 通过观察时间序列的自相关函数 (ACF) 和偏自相关函数 (PACF) 图来确定合适的自回归阶数 p 和移动平均阶数 q 。通过观察 ACF 图, 可以识别时间序列中是否存在任何周期性。PACF 图显示了时间序列与其自身滞后版本之间的部分相关性, 消除了中间滞后阶数的影响。通过 PACF 图可以确定时间序列的滞后阶数, 以建立合适的自回归模型。

(4) 把通过对 ACF、PACF 图像进行观察确定的模型代入检验, 以验证哪一个模型的拟合效果更好。

3 数据及结果

首先, 本文将 1984-2012 纽约证券交易所美股收盘价格作为研究对象, 画出其时间序列图像 (如图 1):

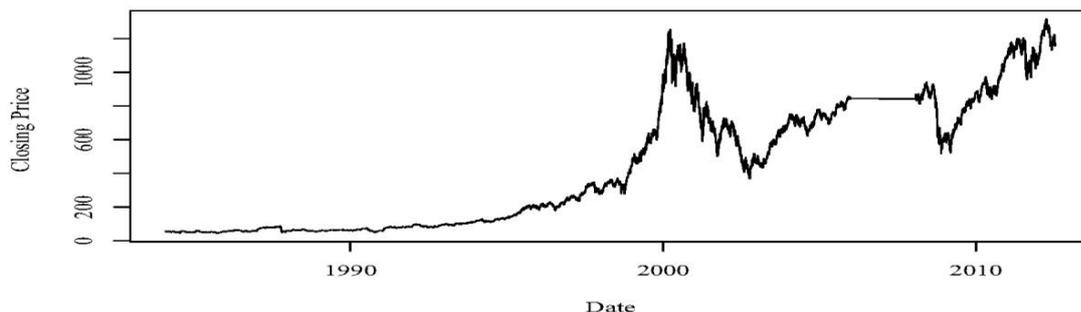


图 1 1984-2012 纽约证券交易所美股收盘价格时间序列图像

由图像特征可以看出, 股票收盘价格的时间序列季节性并不明显, 在 1984 到大概 1997 年之间, 股票价格的波动较为平缓, 一直维持在大约 100 美元左右的水平, 且有一定程度的缓慢增长。这一时期是里根政府时代, 里根政府的减税政策和鼓励企业竞争的措施带动了企业盈利的增长, 同时也推动了股市的上涨。此外, 技术创新和信息技术行业的发展也开始崭露头角, 促进了部分科技公司股价的上涨。在大约 1997 到 2000 年间股票收盘价格经历了迅速增长, 在大约 2000 年间达到第一个峰值, 大约为 1200 美元左右。结合时代背景分析, 这一时期被称为“互联网革命”时期。许多互联网公司如谷歌、亚马逊和雅虎等崛起, 带动了科技股票的疯狂涨势。投资者炒作互联网公司的股票, 使得一些公司的市值飙升。而直到 2000 年, 互联网泡沫的爆发使股市遭受巨大冲击。9.11 恐怖袭击事件导致经济更加不稳定, 加之房地产市场泡沫破裂和金融市场的动荡, 美国陷入了经济衰退。这段时期股市价格波动大, 整体呈现下跌趋势, 在 2000 至大约 2004 年间迅速下降至大约 400 美元, 也就是下降至 2000 年后股票价格的最低值。2004 年之后股票价格又以较快的速度上

涨，在约 2006 年达到 800 元，接着股票价格在大概 2006-2007 年间保持稳定，然后又迎来一段波动，在大约 2009 年间下降到第二个“波谷”，此时股票价格大约为 500 元，最后上涨达到约 1200 元左右。这一时期是金融危机后的复苏阶段，美国政府实施了一系列刺激措施，包括降息和量化宽松政策。这些政策刺激了股市的复苏，企业盈利开始逐步增长。同时，科技行业再次成为股市的引领者，一些新兴科技公司的股价大幅上涨，带动了整个股票市场的价格上升。由此看出，股票收盘价格的时间序列的波动并不具有规律性，受时代背景、政府政策等多方面影响。

接着，在对数据进行平稳化处理，还需要判断该序列是否为随机序列，因为在平稳序列中，纯随机序列并没有建模的价值。判断过程为：先假设序列纯随机，观察 p 值是否满足小于 0.01，若小于 0.01，则表示检验结果拒绝了序列纯随机的原假设，也就是说该序列是一个能够建模分析的序列。把所需数据一一进行检验后，结果发现，在长期的时间区间中，数据都是平稳的，p 值小于 0.01。在验证成功数据能够被建模之后，能够求出该观察值序列的样本自相关系数（Autocorrelation Function，下文简称为 ACF）和样本偏自相关系数（Partial Autocorrelation Function，下文简称为 PACF）的值并画出其图像。

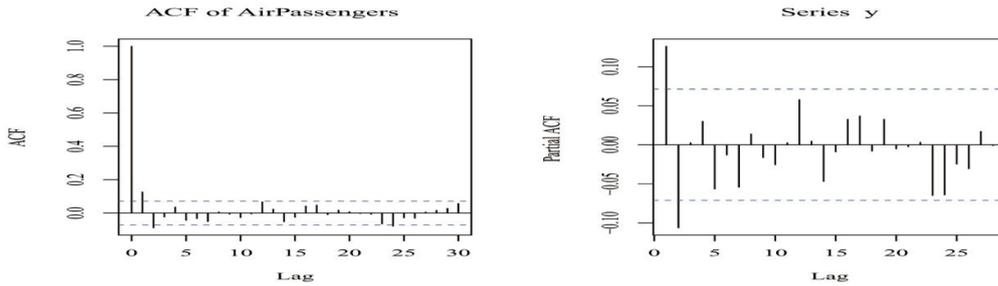


图2 ACF图与PACF图

从图像结果得，自相关图为二阶截尾，偏自相关图为拖尾。接着需要确定模型是否包含差分项。对 ARIMA (p, d, q) 模型而言，p 表示自回归项的阶数，d 表示差分次数，q 表示移动平均项的阶数。在 ARIMA 模型中，差分是通过对时间序列进行减法运算来减少数据的趋势和季节性，使其更加平稳。通过对 acf 图与 pacf 图形特征的观察可得，arima (2, 1, 0) 与 arima (0, 1, 2) 模型可能为能够拟合的模型，因此把这两个模型分别代入检验，以验证哪一个模型的拟合效果更好。首先把 p=2, d=1, q=0 代入上文介绍过的 arima 公式后，得到 ARIMA (2, 1, 0) 的模型形式为：

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t - \varepsilon_{t-1}$$

同理可得，ARIMA (0, 1, 2) 的模型形式为：

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

接着对上述两模型分别进行检验，由检验结果得：

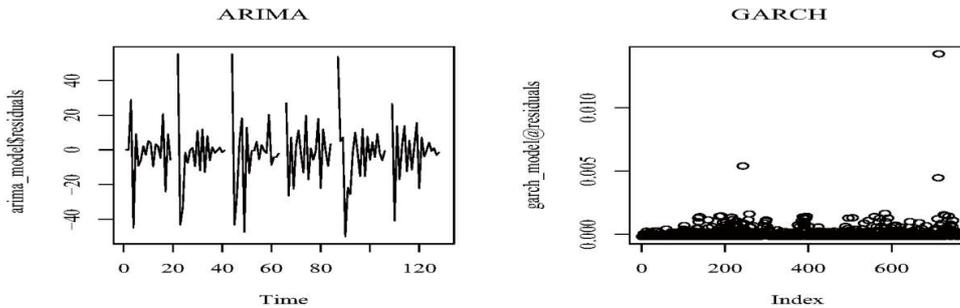


图3 残差检验图

在完成对整个时间序列的分析后，再分别截取两个具有代表性的时间段进行详细分析。首先选择较为平稳的一段时间序列（1984-1990）。

同样，在画出 acf 与 pacf 图像后，

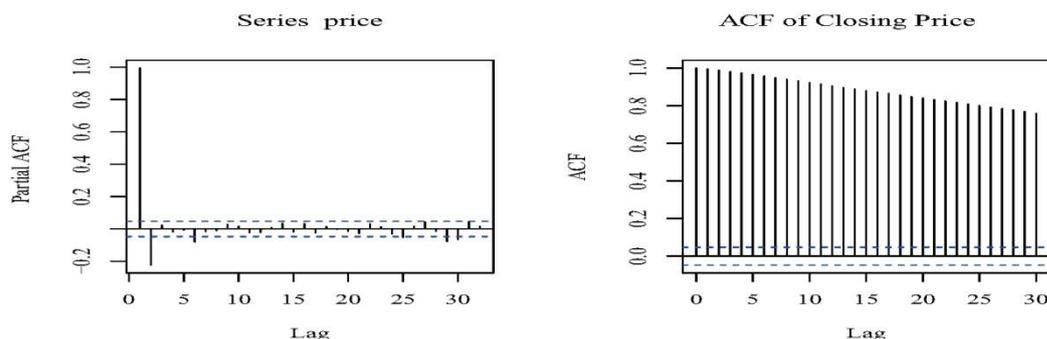


图 4 较平稳区间的 acf 与 pacf 图

通过对其图像特征的观察，分别对其进行阶数为 1 的 AR 模型与阶数为 2 的 AR 模型检验，并观察哪一个模型的检验结果更符合需求，能通过显著性检验且具有更好的拟合效果。由检验结果得，阶数为 2 的 AR 模型显著性检验 p 值等于 0.4177，大于 0.05，通过显著性检验，所以这段较平稳的时间序列能够被二阶 AR 模型拟合。

在对整个时间序列中较为平稳的一段进行模型拟合后，本文再选取波动较为剧烈的一段（1998-2003）进行详细分析：与上文类似，先画出 acf 与 pacf 的图像：

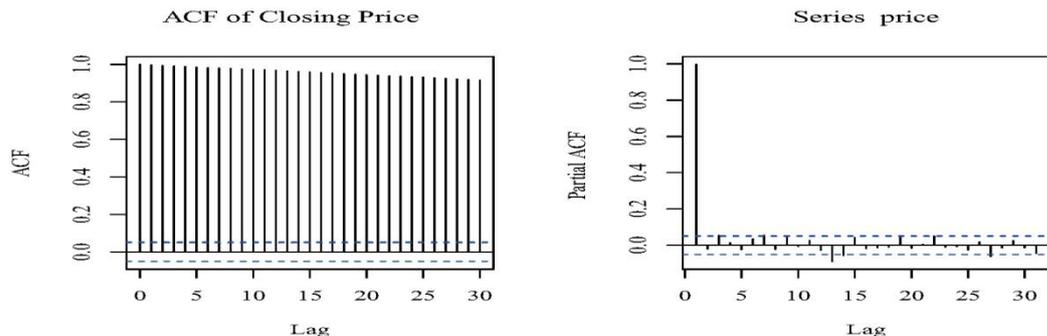


图 5 波动较剧烈区间的 acf 与 pacf 图

通过对其图像特征的观察，分别对其进行阶数为 1 的二阶 MA 模型、阶数为 2 的 MA 模型以及阶数为 3 的 MA 模型检验，来观察是否存在一个对该段时间序列拟合效果较好的模型。由检验结果得，阶数为 2 的 MA 模型显著性检验 p 值等于 0.06133，大于 0.05，通过显著性检验，所以该段波动较剧烈的时间序列能够被二阶 MA 模型拟合。

由此可得，对股票价格的波动情况，总可以找到适合的时间序列模型对其进行拟合，因此通过建立适合的模型，投资者可以利用过去的股票价格数据来预测未来价格走势，从而识别和量化风险，找到相应的风险管理措施应对风险。不仅如此，投资者还可以更好的确定买入与卖出时机，根据模型预测调整交易策略，确定最优的资产配置方案，实现收益最大化。

4 结论与展望

本文主要研究了自回归移动平均模型在股票收盘价格时间序列分析中的应用，本文主要贡献聚焦于长时间跨度的股票走势分析，通过对 1984-2012 纽约证券交易所美股收盘价格的研究，分别对其总的时间序列

以及两个具有代表性的分段进行了三种模型的成功拟合,验证了股票收盘价格可以通过该模型进行较好的预测,对投资者的投资决策以及金融市场的维稳具有一定价值。然而股票的波动趋势受多种因素影响,本文仅通过历史数据对美股收盘价格进行了分析,未考虑到股息收益率、资金流向等可以影响股票价格的指标。如何得到更加精确的预测结果还需未来进一步的研究。

参考文献

- [1] 程孟菲;高淑萍;.基于深度迁移学习的多尺度股票预测[J].计算机工程与应用,2022(12).
- [2] Zhang Q, Yang L, Zhou F. Attention enhanced long short-term memory network with multi-source heterogeneous information fusion: An application to BGI Genomics[J]. Information Sciences, 2021, 553:305-330.
- [3] Zhao J, Zeng D, Liang S, et al. Prediction model for stock price trend based on recurrent neural network[J]. Journal of Ambient Intelligence and Humanized Computing, 2021,12(1): 745-753.
- [4] 张艺萧. 基于时间序列的股票价格分析与预测[D].华北电力大学(北京),2020.
- [5] 李奋华,赵润林. 一种基于时间序列分析的股票走势预测模型[J].现代计算机(专业版), 2016, (20): 14-17.
- [6] 杨智勇,叶玉玺,周瑜. 基于 BiLSTM-SA-TCN 时间序列模型在股票预测中的应用[J]. 南京信息工程大学学报(自然科学版), 2023, 15 (06): 643-651.
- [7] 高雅婷.时间序列趋势预测方法研究及其在股票市场的应用[D].山东工商学院, 2022.
- [8] 陈登建,杜飞霞,夏换. 基于 ARIMA 与 SVR 滚动残差模型组合的股票预测[J]. 计算机时代, 2022, (05): 76-81.
- [9] Liang M , Wu S , Wang X ,et al.A stock time series forecasting approach incorporating candlestick patterns and sequence similarity[J].Expert Systems with Application, 2022.
- [10] Pin Lv ,Qinjuan Wu,Jia Xu,et al.Stock Index Prediction Based on Time Series Decomposition and Hybrid Model[J].Entropy,2022
- [11] Horák Jakub;Krulický Tomáš.Comparison of exponential time series alignment and time series alignment using artificial neural networks by example of prediction of future development of stock prices of a specific company.[J].SHS Web of Conferences.Volume 61, Issue .2019. PP 01006.

版权声明: ©2024 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



OPEN ACCESS