

基于滚动ARIMA预测模型对2023年6月 比特币收盘价的预测

崔博涵

成都理工大学数理学院, 四川 成都

收稿日期: 2023年5月21日; 录用日期: 2023年6月14日; 发布日期: 2023年6月21日

摘 要

近年来, 比特币已经成为了最流行的加密货币之一, 它的价格波动经常被关注和研究, 目前比特币的市场价值非常高, 已经达到了历史新高, 在全球范围内引起了广泛的兴趣和关注, 它已经成为了数字资产领域的重要一员。本文通过研究时间序列模型在预测中的应用, 提出了滚动ARIMA预测模型, 使用MASE, MAE, RMSE比较了三次指数平滑模型和滚动ARIMA模型的预测效果, 最后预测了2023年6月的比特币收盘价。研究表明, 在短期预测方面, 滚动ARIMA模型和三次指数平滑预测模型效果大致相同, 但是滚动ARIMA模型具有较好的方向控制和误差控制效果, 最终得出滚动ARIMA模型适合长期预测的结论。结果显示, 在2023年6月, 比特币收盘价呈现波动趋势, 并且波动趋于平缓。

关键词

比特币收盘价, 滚动ARIMA预测模型, 三次指数平滑模型, MATLAB

Prediction of the Closing Price of Bitcoin in June 2023 Based on the Rolling ARIMA Forecast Model

Bohan Cui

College of Mathematics and Physics, Chengdu University of Technology, Chengdu Sichuan

Received: May 21st, 2023; accepted: Jun. 14th, 2023; published: Jun. 21st, 2023

Abstract

In recent years, Bitcoin has become one of the most popular cryptocurrencies, its price fluctuations are often paid attention to and studied, the current market value of Bitcoin is very high, has reached a record high, has attracted widespread interest and attention worldwide, it has become an important member of the digital asset field. By studying the application of time series model in forecasting, this paper proposes a rolling ARIMA forecasting model, compares the prediction effect of the three exponential smoothing model and the rolling ARIMA model using MASE, MAE, RMSE, DAR and finally predicts the closing price of Bitcoin in June 2023. The results show that in short-term prediction, the rolling ARIMA model and the cubic exponential smoothing prediction model have roughly the same effect, but the rolling ARIMA model has better direction control and error control effect, and finally concludes that the rolling ARIMA model is suitable for long-term prediction. The results showed that in June 2023, the Bitcoin closing price showed a volatile trend and the volatility flattened.

Keywords

Bitcoin Closing Price, Rolling ARIMA, Triple Exponential Smoothing, MATLAB

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

近年来, 比特币投资研究不断增加, [1]指出了目前比特币研究主要包括利用量化投资模型对比特币投资模型的构建, 利用时间序列模型和多因子模型预测比特币的涨跌, 利用机器学习模型等结合大数据进行量化投资。[2]建立了 GARCH 模型对比特币波动率进行预测。[3]通过基于神经网络的方法预测比特币价格变化。

关于时间序列预测, 很多学者提出了各种预测模型, 其中包括指数平滑预测[4] [5]、三次指数平滑预测[6] [7]、动态指数平滑预测[8] [9]、ES-GA 预测[10]、ARIMA 预测[4], [11] [12]、滚动 ARIMA 预测[13]、SVM 预测[14]以及 LSTM 预测[11] [12]等。但是, 目前很少有关于滚动 ARIMA 模型的比特币收盘价预测相关文献。

目前的研究中, 指数平滑模型广泛应用于时间序列分析中, 三次指数平滑模型可以很好地保存时间序列数据的趋势和季节性信息。文献[7]研究表明, 三次指数平滑模型能够根据时序的变化, 清晰地显示出时序的变化趋势, 对波动范围较大且呈非线性变化规律的数据具有很强的适用性, 具有预测可信度较高、基础数据周期要求短、操作便捷且使用方便等优点。关于 ARIMA 的预测模型, 大多数都是整体预测或者结合深度学习方法预测, 在预测阶数比较大的时候, 误差会非常大。文献[13]提出了滚动 ARIMA 预测模型, 滚动预测模型的主要思想是, 每次预测的时候, 只预测一期数据, 下次预测的时候把上期预测的数据加到训练集, 依此类推, 这极大地减少了时间序列自相关性降低所带来的误差。但是, 目前已知的滚动 ARIMA 预测模型的文献中, 训练集中数据的数量是随着预测次数而增加的, 这可能导致过拟合而降低泛化能力。

本文提出了一种新的滚动 ARIMA 预测模型, 即每次预测单期数据后, 得到的数据加入到训练集,

同时，训练集的第一次数据被移出训练集，这样就保证了每次预测时，训练集中数据的数量保持不变。本文主要通过滚动 ARIMA 预测和动态指数平滑方法，在比较测试集上的相对误差的同时，预测 2023 年 6 月的比特币收盘价。其中第二节讨论了滚动 ARIMA 和动态指数平滑的方法，第三节比较两种方法的效果，最后一节是本研究的结论。

2. 研究方法

2.1. 滚动 ARIMA 模型

2.1.1. 传统的 ARIMA 预测模型[4]

ARIMA 模型是由自回归(AR)模型和移动平均(MA)模型组成的时间序列预测模型[11]。它与 ARMA 模型属于同一自回归模型，但对数据的要求却大不相同。ARMA 模型适用于稳态时间序列的数据，而 ARIMA 模型更适合差分后稳态时间序列的数据[11]，ARIMA 模型全称为差分自回归移动平均模型，广泛适用于非平稳时间序列，令我们的模型为 $ARIMA(p, d, q)$ ，如果 $\{X_t\}$ 是时间序列观测值， μ 是序列的均值常数， ε_t 是序列的误差， φ_i 是自回归系数， θ_j 是移动平均系数， L^d 是 d 阶差分算子， $L(X_t) = X_t - X_{t-1}$ ， $L^d(X_t) = L(L^{d-1}(X_t))$ ，这里 $i=1, 2, \dots, p$ ； $j=1, 2, \dots, q$ ，则 ARIMA 模型的表达式为(1)

$$L^d(X_t) = \mu + \sum_{i=1}^p \varphi_i L^d(X_{t-i}) + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{1}$$

传统的 ARIMA 预测模型的思路如下：1) ADF 检验，首先对序列进行 ADF 平稳性检验，如果序列不平稳，则对序列进行差分，直到序列平稳为止。若对序列的差分阶数为 d 时序列通过了平稳性检验，则可以确定该模型的参数 d 。我们对 2016 年 4 月到 2023 年 5 月的比特币收盘价进行了 pp 检验，如表 1 所示。由于 $p < 0.05$ ，可以断言，比特币收盘价数据是不平稳的，需要进行差分平稳化。2) 白噪声检验[15]，序列平稳后，对序列进行白噪声检验，即检验时间序列是否为纯随机的，如果这个时间序列具有纯随机性，则认为序列中不包含任何值得提取的有用信息，即停止对序列进行分析，检验白噪声可用 Q 统计量或 LB 统计量。本文中 2016 年 5 月到 2023 年 5 月比特币收盘价数据白噪声检验的 p 值小于 0.05，所以可认为序列为非白噪声。3) 模型定阶，根据 AIC 或 BIC 信息准则[16]，即最小化信息量准则或贝叶斯信息准则，选取最小的 AIC 或 BIC 所对应的 p 和 q ，在这里， $ARIMA(p,d,q)$ 模型定阶已经完成。4) 参数估计，定阶完成后，要对 ARIMA 模型的全部未知参数 φ_i, θ_j 进行估计，文献[17]提出了 PHS 算法进行参数估计。Matlab 可用直接实现参数估计。5) 预测，参数估计完成后，传统的 ARIMA 模型可以预测后面多期的数据。为提高模型准确性，我们在 2.1.2 引入了滚动 ARIMA 预测模型。6) 异常值处理，由于比特币收盘价具有非负性，所以如果预测到的某一期数据为负值，则令这一次预测的数据为 0。

Table 1. Nonparametric test of the Bitcoin closing price series

表 1. 比特币收盘价序列的非参数检验

p-Value (pp test)	p-Value (LB, Lag = 6)	p-Value (LB, Lag = 12)	p-Value (LB, Lag = 18)
0.4433	0.0000	0.0000	0.0000

传统的 ARIMA 预测模型如图 1。

也就是说，对一个时间序列建立一个整体的模型，利用这个模型预测后面的数据，如果利用单个模型预测长期数据的话，预测结果会接近一条直线，效果会变得很差。但是，如果每次只预测一期数据，然后把新预测的数据加入训练集，去预测下一期数据，这样的话预测精度会提高很多。因此，我们考虑滚动 ARIMA 预测模型。

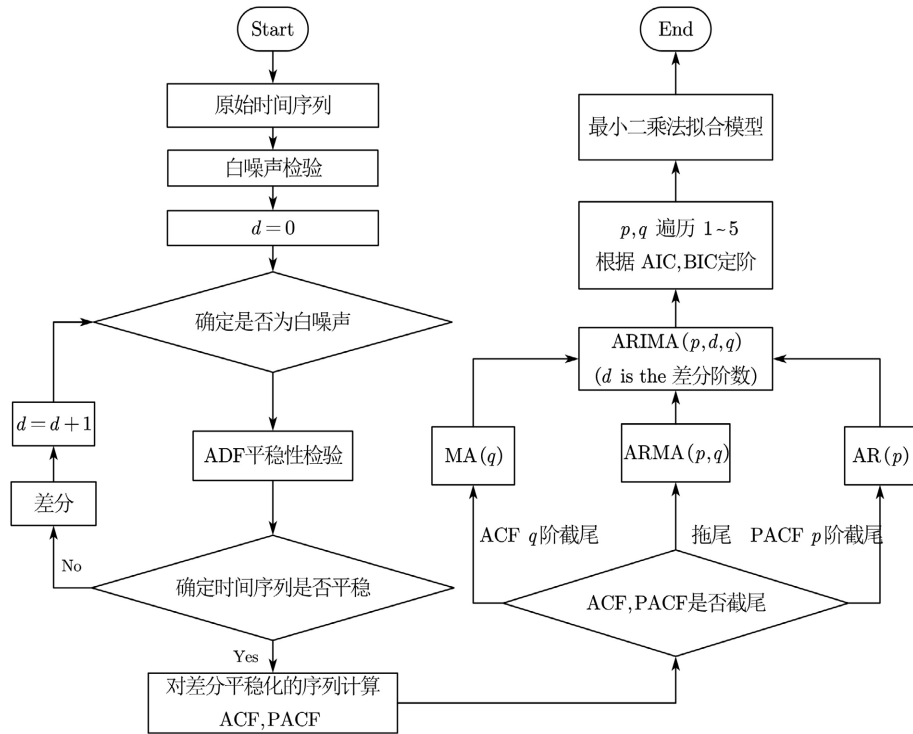


Figure 1. Flowchart of ARIMA
图 1. ARIMA 模型流程图

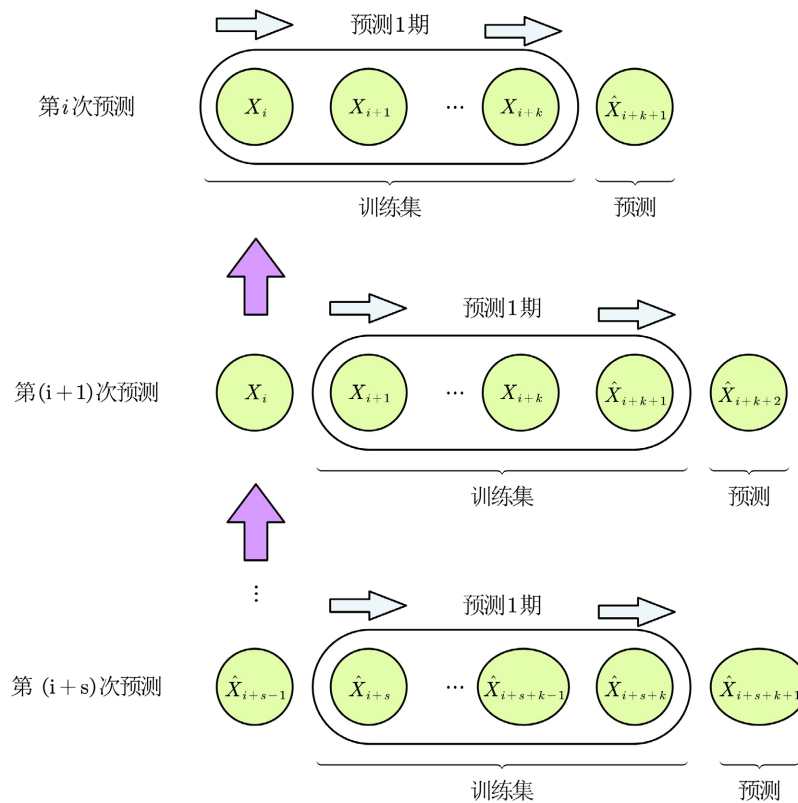


Figure 2. Diagram of rolling ARIMA
图 2. 滚动 ARIMA 模型示意图

2.1.2. 滚动 ARIMA 预测模型

滚动 ARIMA 预测模型的主要思路是：提前确定训练集数据的数量 k ，每次都是单步预测，即每次只预测后面一期数据的值，然后把这个预测的值加入到训练集，再把训练集的最早期的数据移出训练集，这样可以确保训练集的数据数量恒定不变，以减少过拟合并且提高泛化能力。例如，第一次预测，训练集为 $T_1 = \{X_1, X_2, \dots, X_k\}$ ，利用这些训练集预测的下一期的数据是 \hat{X}_{k+1} ，我们把 \hat{X}_{k+1} 加入到训练集，再把训练集的第一个数据移出，第二次预测，训练集为 $T_2 = \{X_2, \dots, X_k, \hat{X}_{k+1}\}$ ，依此类推，预测的数据为 $\{\hat{X}_{k+1}, \hat{X}_{k+2}, \dots, \hat{X}_{k+n}\}$ ，其中 n 为预测总期数。滚动 ARIMA 的原理如图 2。传统 ARIMA 模型的时间复杂度是 $O(1)$ ，滚动 ARIMA 预测模型的时间复杂度是 $O(n)$ [13]，滚动预测在以提高时间复杂度为代价的前提下，提高了预测精度。

2.2. 三次指数平滑模型

三次指数平滑模型是由一次指数平滑模型发展而来的。指数平滑法是一种时间序列分析和预测算法，包括一次、二次、三次指数平滑[18]。一次指数平滑法是对当前周期的观测值和预测值赋予不同的权重，从而得到下一个周期的预测值的方法。但是一阶指数平滑不能处理具有长期趋势或季节性影响的时间序列[8]。二次指数平滑法是将一次平滑的结果再次平滑为实际值，从而保留趋势信息。但二次指数平滑法不能处理具有季节效应的时间序列。三次指数平滑在二次指数平滑的基础上保留了季节信息，可以预测具有季节效应的时间序列。文献[7]提到三次指数平滑的算法如下：设 $\{X_t\}$ 是第 t 天的比特币收盘价，则可以构建三次指数平滑预测模型(2)

$$\begin{cases} Y_{t+T} = a_t + b_t T + c_t T^2 \\ a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \frac{\alpha}{2(1-\alpha)^2} \left[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)} \right] \\ c_t = \frac{\alpha^2}{2(1-\alpha)^2} \left[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)} \right] \\ S_t^{(1)} = \alpha X_t + (1-\alpha)S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \end{cases} \quad (2)$$

这里 Y_{t+T} 表示预测的第 T 天的比特币收盘价， T 为超前预测的期数。 a_t, b_t, c_t 是第 t 天的预测系数， $S_t^{(1)}, S_t^{(2)}, S_t^{(3)}$ 分别为第 t 天的三次指数平滑值， α 是平滑系数。

我们首先对未来 31 天的比特币收盘价进行预测，我们使用的工具是 Matlab R2021b。预测的对比图如图 3，图 4。

3. 模型之间的比较

3.1. 比较模型所使用的统计量

为了对比滚动 ARIMA 预测模型和三次指数平滑预测模型的效果，我们对比如下的统计量：相对百分误差(MAPE)，平均绝对值误差(MAE)，均方误差(RMSE) [19]，符号正确率(DAR) [20]

MAE：平均绝对误差，是预测值和实际值之间的绝对误差算术平均值，如(3)

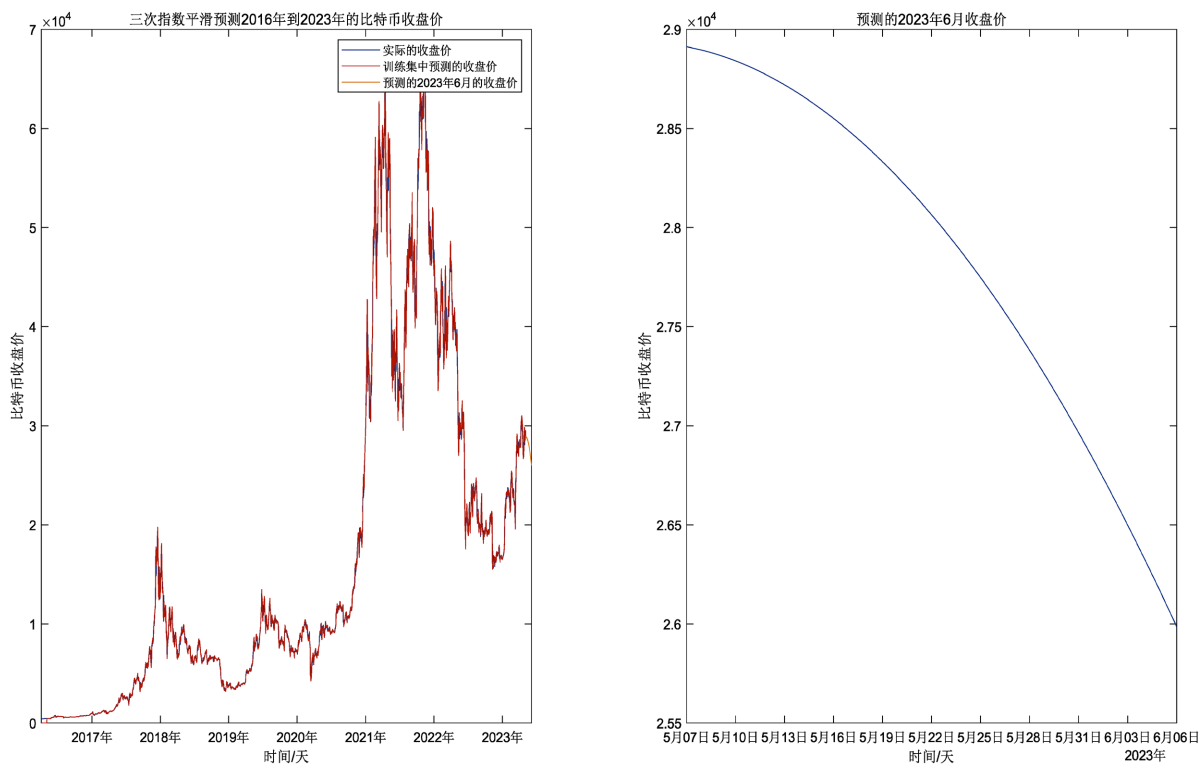


Figure 3. Bitcoin closing price prediction in 2023.06 by cubic exponential smooth

图 3. 三次指数平滑的比特币收盘价预测

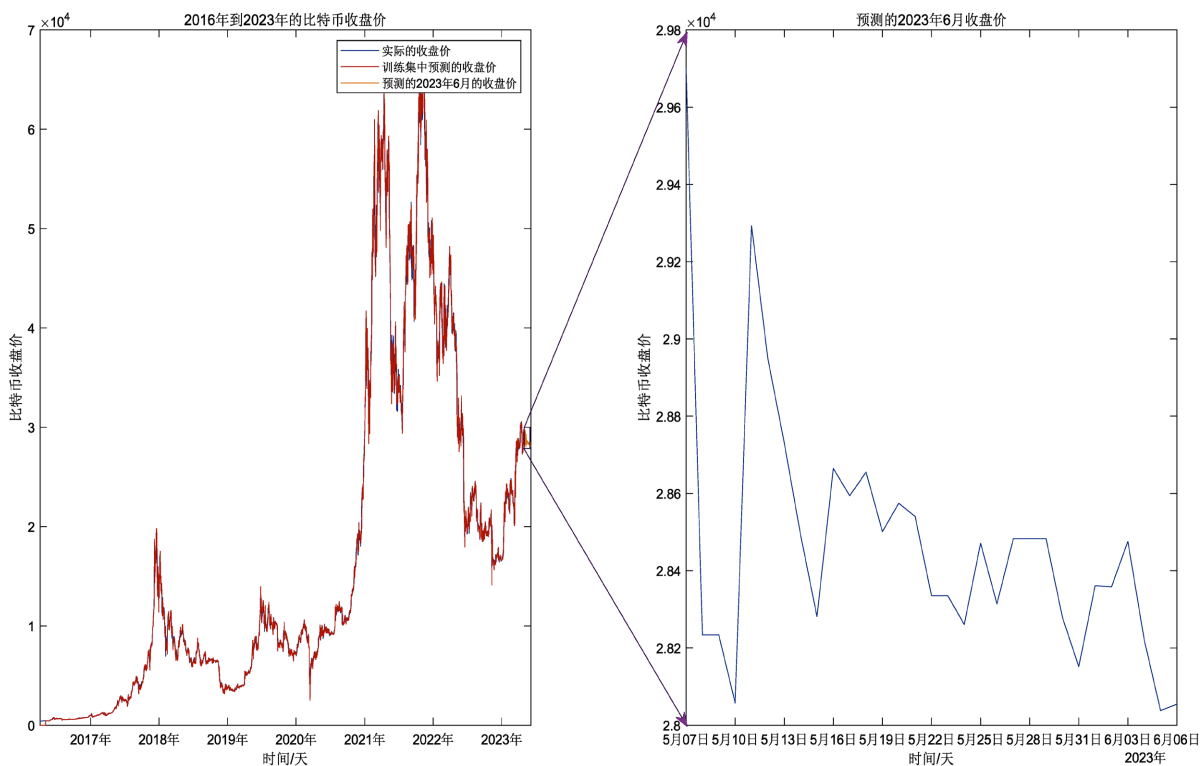


Figure 4. Bitcoin closing price prediction in 2023.06 by rolling ARIMA

图 4. 滚动 ARIMA 的 2023 年 6 月比特币收盘价预测

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - F_t| \tag{3}$$

MAPE: 是平均绝对百分比误差, 是预测值与实际值之间的绝对误差与实际值比值的算术平均值, 定义如

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{A_t - F_t}{A_t} \times 100\% \tag{4}$$

RMSE: 均方根误差, 是对预测值与实际值之差的平方和的平均值进行开方得到的, 定义如(5)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n |A_t - F_t|^2} \tag{5}$$

DAR: 符号正确率, 是预测的符号正确的累计次数与总预测次数的比值, 通常 DAR 是评价金融相关预测最重要的统计量, DAR 的定义如(6)。

$$DAR = \frac{1}{n} \sum_{t=1}^n z_t \tag{6}$$

其中, z_t 通过(7)式定义:

$$z_t = \begin{cases} 1, & (F_{t+1} - A_t)(A_{t+1} - A_t) \geq 0 \\ 0, & (F_{t+1} - A_t)(A_{t+1} - A_t) < 0 \end{cases} \tag{7}$$

在这里, A_t 表示第 t 年比特币收盘价的实际值, F_t 表示第 t 年比特币收盘价的观测值。

3.2. 基于滚动 ARIMA 和三次指数平滑对比特币收盘价的预测比较

对于滚动 ARIMA 模型, 本文采用 2016 年 5 月到 2023 年 5 月的比特币收盘价进行实验, 用前 30 天 (2016.5~2016.6) 的比特币收盘价作为初始数据, 生成滚动 ARIMA 预测模型, 测试集为 2016.6~2023.5 年的比特币收盘价, 比特币收盘价的滚动 ARIMA 预测结果如图 3, 图 4; 对于三次指数平滑模型, 训练集和测试集同滚动 ARIMA 模型, 我们选取 $\alpha = 0.3$, 使用滚动预测的方法, 即 $T = 1$, 预测 2016.6~2023.5 比特币收盘价, 预测结果在图 3, 图 4 通过对比测试集的 MAPE, MAE, RMSE, DAR 以量化模型效果, 求解结果如表 2。

Table 2. Comparison of performance between models
表 2. 模型效果对比

	MAPE	MAE	RMSE	DAR
指数平滑模型	4.3445%	540.7107	1019.6	49.75%
滚动 ARIMA 模型	4.3316%	538.2423	1045.1	57.9%

3.3. 基于滚动 ARIMA 预测模型对未来 31 天比特币收盘价的预测

由于滚动 ARIMA 模型在测试集中的误差表现良好, 即在测试集中, 符号正确率 DAR 为 57.9%, 但是对于三次指数平滑模型, 测试集中, 符号正确率 DAR 小于 50%, 这说明, 滚动 ARIMA 预测模型具有较好的方向预测以及误差控制效果。由于长期预测时, 误差会越来越大。因此, 在预测 31 天后比特币收盘价时, 我们选用方向预测和误差控制较好的滚动 ARIMA 模型, 预测结果在图 3, 图 4。

4. 结论

时间序列分析是基于历史观测数据对未来数据进行预测的一种方法, 广泛应用于各个学科。为了预

测 31 天后的比特币收盘价, 我们先后考虑指数平滑模型, ARIMA 模型, 在现有的 ARIMA 模型上使用了改进模型, 即滚动 ARIMA 模型, 在提高了时间复杂度的前提下增加了预测的准确度。

我们把 2016 年 6 月到 2023 年 5 月的比特币收盘价作为测试集, 对比特币收盘价预测效果进行了测试, 用于比较两种模型的精度, 本研究结果显示: 整体上两种模型的 MAPE 都比较小, 效果相当。对于滚动 ARIMA 模型, 方向准确率 DAR = 57%, 但是对于三次指数平滑模型, 方向正确率不足 50%。结果表明, 滚动 ARIMA 预测模型的方向预测和误差控制效果较好。由于长期预测时, 误差会越来越大。因此, 在预测 31 天后比特币收盘价时, 我们选用预测方向和误差控制较好的滚动 ARIMA 模型。

基于滚动 ARIMA 预测模型, 我们预测了未来 31 天的比特币收盘价, 结果显示, 在 2023 年 6 月, 比特币收盘价呈现波动趋势, 并且波动趋于平缓。

参考文献

- [1] 马基远. 数字货币量化投资模型[D]: [硕士学位论文]. 北京: 首都经济贸易大学, 2021.
- [2] 王倩, 杜卓雅. 比特币与黄金避险功能的差异研究——基于 VAR-BEKK-GARCH 模型[J]. 管理科学, 2022, 35(2): 134-146.
- [3] 张杰夫. 基于机器学习的比特币价格预测及影响因素研究[D]: [硕士学位论文]. 成都: 西南财经大学, 2021.
- [4] 任嘉豪, 徐洁, 杨海燕. ARIMA 及 Holt-Winters 指数平滑模型在河南省肺结核流行趋势预测中的应用[J]. 郑州大学学报(医学版), 2022, 57(6): 756-760.
- [5] 贡文伟, 黄晶. 基于灰色理论与指数平滑法的需求预测综合模型[J]. 统计与决策, 2017(1): 72-76.
- [6] 韩芹芹, 李凡, 王涛, 张守斌. 基于三次指数平滑法的乌鲁木齐市声环境质量污染预测及应用[J]. 环境科学导刊, 2022, 41(5): 84-89. <https://doi.org/10.13623/j.cnki.hkdk.2022.05.012>
- [7] 曹志成, 刘伊生, 董继伟. 基于三次指数平滑法的铁路旅客发送量预测研究[J]. 铁道运输与经济, 2018, 40(11): 49-53+115. <https://doi.org/10.16668/j.cnki.issn.1003-1421.2018.11.10>
- [8] 黄伟建, 张一帆, 黄远. 动态三次指数平滑法的火电厂发电量预测研究[J]. 现代电子技术, 2020, 43(17): 147-150+154.
- [9] 孟利民, 徐杨. 基于动态指数平滑预测的负载均衡算法[J]. 浙江工业大学学报, 2016, 44(4): 379-382.
- [10] 张英贵, 杨蕙瑜, 雷定猷. 基于组合输入 ES-GA-BP 的中欧班列货运量预测[J]. 深圳大学学报(理工版), 2022, 39(2): 168-176.
- [11] 沈露露, 梁嘉乐, 周雯. 基于 ARIMA-LSTM 的能量预测算法[J/OL]. 无线电通信技术: 1-8. <http://kns.cnki.net/kcms/detail/13.1099.TN.20230106.0919.004.html>, 2023-01-16.
- [12] 次必聪, 张品一. 基于 ARIMA-LSTM 模型的金融时间序列预测[J]. 统计与决策, 2022, 38(11): 145-149.
- [13] 林在超. 基于滚动 ARIMA 的农作物价格预测方法应用[J]. 现代计算机, 2021, 27(35): 80-83.
- [14] Lee, P.H. (2022) ARIMA and SVM Forecasting in the US Paper Waste. *2nd International Conference on Applied Mathematics, Modelling, and Intelligent Computing (CAMMIC 2022)*, Kunming, 25-27 March 2022, 122595W.
- [15] 李为东, 李莉, 徐岩. 基于时间序列分析的北京地区 PM_{2.5} 浓度研究[J]. 运筹学学报, 2018, 22(2): 115-126.
- [16] 袁路妍, 王占宏. 基于 ARIMA 的入境旅游月度过夜人次预测[J]. 微型电脑应用, 2020, 36(4): 7-9.
- [17] 单锐, 王国芳, 黄威, 刘文, 王美霞. 基于改进谱共轭梯度思想的 ARIMA 模型参数估计优化法[J]. 兰州理工大学学报, 2018, 44(4): 152-156.
- [18] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2016.
- [19] 郝广涛, 林清华, 李晓梅. 超短期负荷预测中指数平滑法平滑系数的确定方法[J]. 莆田学院学报, 2020, 27(5): 80-86.
- [20] 龙奥明, 毕秀春, 张曙光. 基于 LSTM 神经网络的黑色金属期货套利策略模型[J]. 中国科学技术大学学报, 2018, 48(2): 125-132.