

# 基于ARIMA模型的股票价格预测分析

罗 露

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2022年11月26日; 录用日期: 2022年12月21日; 发布日期: 2022年12月30日

## 摘 要

本文立足于对时间序列数据的研究分析, 以深粮控股的股价数据为实验对象, 采用一种基于差分自回归移动平均(ARIMA)模型对其未来几天的股票价格进行预测。用python和Eviews对采集的股票价格数据进行平稳性检验、白噪声检验、模型定阶、残差检验等步骤。建立了有效预测股票价格的ARIMA模型, 同时结合真实值, 对模型的有效性进行检验。结果表明, 该方法能有效提取原始数据中心的信息, 对股票价格预测效果较好。

## 关键词

时间序列, 股票价格, ARIMA模型, 预测

# Stock Price Forecasting Analysis Based on ARIMA Model

Lu Luo

School of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Nov. 26<sup>th</sup>, 2022; accepted: Dec. 21<sup>st</sup>, 2022; published: Dec. 30<sup>th</sup>, 2022

## Abstract

Based on the research and analysis of time series data, this paper adopts a differential autoregressive moving average (ARIMA) model to predict the stock price of SZG Holdings in the coming days based on its stock price data as the experimental object. Firstly, the collected stock price data were tested for smoothness, white noise test, model sizing and residual test by using python and Eviews, and then an ARIMA model was established to effectively predict the stock price, while the validity of the model was tested by combining the true values. The results show that the method can effectively extract information from the center of the original data and has a good effect on stock price prediction.

## Keywords

Time Series, Stock Prices, ARIMA Model, Forecasting

Copyright © 2022 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

金融市场在国家经济体系中扮演着至关重要的角色，金融市场的表现反映着国家经济发展状态，挖掘金融市场变化的潜在规律并准确的判断金融市场的发展趋势，有利于相关从业者制定更完善的策略，达到降低风险，提高效益的目的[1]。

一直以来，股票在金融市场中有着举足轻重的地位，而股票价格往往对时间因素非常敏感[2]，所以建立时间序列模型能有效地对股票价格进行预测。本文选取深粮控股的收盘价历史数据，采用 ARIMA 模型，对未来几天的收盘价进行预测。

## 2. ARIMA 模型原理

ARIMA 模型即差分移动平均自回归(Autoregressive Integrated Moving Average)模型，主要被用于拟合具有平稳属性的时间序列，或者是可以转换为具有平稳属性的时间序列，是目前广泛为学者应用的时间序列拟合方法[3]。其中心思想是通过差分运算法使得不平稳的原始序列趋于平稳，其中  $d$  为差分运算的总次数。基本原理如下：

$$\begin{aligned} \varphi(B)\nabla^d y_t &= \theta(B)\varepsilon_t \\ E(\varepsilon_t) &= 0, \text{var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(y_s \varepsilon_t) &= 0, \forall s < t. \end{aligned} \quad (1)$$

其中， $\nabla^d = (1-B)^d$ ； $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ 、 $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  分别为平稳可逆 ARIMA( $p, q$ ) 模型的自回归系数多项式和移动平均系数多项式。其中自回归(AR( $p$ ))、移动平均(MA( $q$ ))、自回归移动平均(ARMA( $p, q$ ))模型都是 ARIMA( $p, d, q$ )模型的一个特例。当  $q = d = 0$  且  $p \neq 0$  时，ARIMA( $p, 0, 0$ )即为 AR( $p$ )模型；当  $p = d = 0$  且  $q \neq 0$  时，ARIMA( $0, 0, q$ )即为 MA( $q$ )模型；当  $d = 0$  且  $p \neq 0, q \neq 0$  时，ARIMA( $p, 0, q$ )即为 ARMA( $p, q$ )模型[4]。ARIMA 模型的性质见下表 1：

Table 1. ARIMA( $p, d, q$ ) fixed-order rules

表 1. ARIMA( $p, d, q$ )定阶规则

模型	自相关系数 ACF	偏自相关系数 PACF
ARIMA( $p, d, 0$ )	拖尾	$p$ 阶截尾
ARIMA( $0, d, q$ )	$q$ 阶截尾	拖尾
ARIMA( $p, d, q$ )	拖尾	拖尾

## 3. ARIMA 模型构建

大致来讲，ARIMA 模型构建大约需要几个步骤，如下图 1 所示：

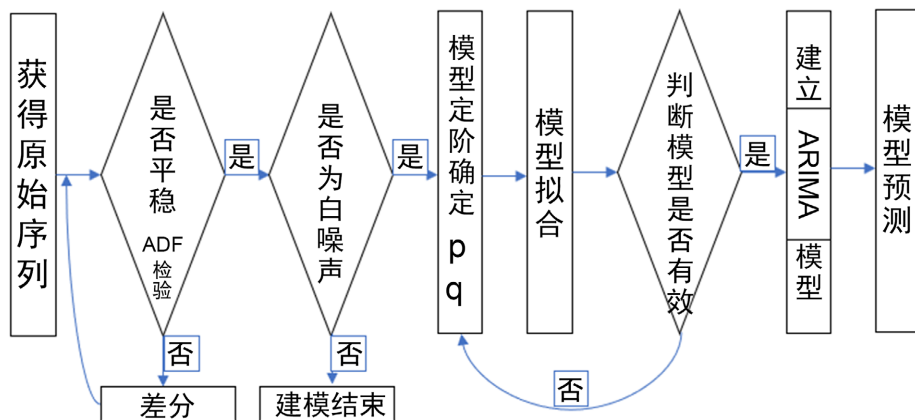


Figure 1. Flow chart of ARIMA model construction

图 1. ARIMA 模型构建流程图

平稳性检验。时间序列的平稳性代表时间序列的统计性质关于时间平移的不变性。对原始股票价格数据进行平稳性检验时，如果数据为非平稳时间序列，需要对序列进行差分运算，直至序列平稳[5]。

白噪声检验(纯随机性检验)。序列的各项数值之间不相关，序列在进行完全无序的随机波动，这样的序列称为纯随机时间序列(白噪声序列)。纯随机时间序列是没有信息可提取的序列，如果序列为白噪声序列，应当停止分析[6]。

模型定阶(模型识别)。结合数据样本时序图和 AIC 信息准则，确定模型自回归系数  $p$  和移动平均系数  $q$  的最优组合[7]。

模型拟合。根据已经确定的最优参数  $p$  和  $q$ ，对平稳非白噪声时间序列进行拟合，构建  $ARIMA(p,d,q)$  模型。

模型检验。对模型进行显著性(有效性)检验，检验模型是否将样本信息进行充分提取。当模型检验为无效模型时，重复步骤三，直至模型有效[8] [9]。

### 3.1. 数据描述

本实验所采集数据均来自 tushare 金融大数据社区，通过 python 连接 API 接口获得。实际选取了深粮控股的收盘价 2021 年 12 月 6 日至 2022 年 7 月 15 日共 147 个数据样本。

### 3.2. 数据预处理

为保证数据的可靠性和有效性，本文数据全部选自 tushare 金融大数据社区，然后对其缺失值和异常值进行了处理，对缺失值全部采用上一个交易日的数据，对异常值作删除处理。

### 3.3. 模型构建

首先，对原始数据进行平稳性检验，观察原序列的时序图如图 2 所示。

观察图 2，可见其为非平稳序列，不过为避免肉眼观察的主观性，再对原始序列进行 ADF 单位根检验。即序列若存在单位根，则不平稳。ADF 单位根检验结果如表 2 所示，由表可知，检验统计量  $adf = -0.002086 > -1.61527 > -1.943012 > -2.580788$ ，且  $P = 0.6504 > 0.05 = \alpha$ ，不能拒绝原假设，故判断原序列非平稳。

ARIMA 模型建模要求必须为平稳的时间序列，因此对原序列进行一阶差分，再对一阶差分后的序列进行 ADF 检验，结果如表 3 所示：



**Figure 2.** Original data time series diagram  
**图 2.** 原始数据时序图

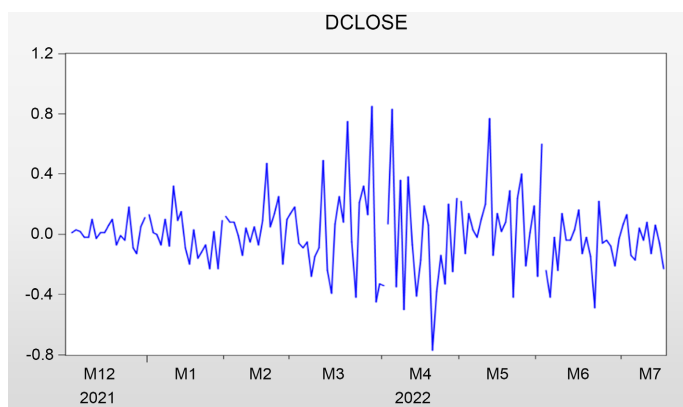
**Table 2.** Results of the original series ADF test  
**表 2.** 原序列 ADF 检验结果

adf	P	显著性水平		
		1%	5%	10%
-0.002086	0.6804	-2.580788	-1.943012	-1.615270

**Table 3.** Results of the ADF test for the series after first-order differencing  
**表 3.** 一阶差分后序列 ADF 检验结果

adf	P	显著性水平		
		1%	5%	10%
-13.18717	0	-2.580788	-1.943012	-1.615270

观察图 3 差分后序列已平稳,由表 3, 检验统计量  $adf = -13.18717 < -2.580788 < -1.943012 < -1.61527$ , 且  $P$  值  $= 0 < 0.05 = \alpha$ , 拒绝原假设, 故判断序列平稳, 可以用于建模。



**Figure 3.** Timing diagram of the sequence after first-order differencing  
**图 3.** 一阶差分后序列时序图

一阶差分后序列平稳，故确定  $d = 1$ ，模型的定阶就是估计自回归阶数  $p$  和移动平均阶数  $q$  的过程。图 4 为差分后序列的自相关和偏自相关图，观察图 4 可知，ACF 和 PACF 均拖尾，判断为 ARMA 模型，初步判断取  $p = q = 5$ 。为进一步确定更合适的  $p$  和  $q$  取值，运用 AIC 信息准则估计不同阶数对模型拟合精度的影响，如表 4 所示，综合结果选择  $p = q = 5$ ，建立 ARIMA(5,1,5)模型。

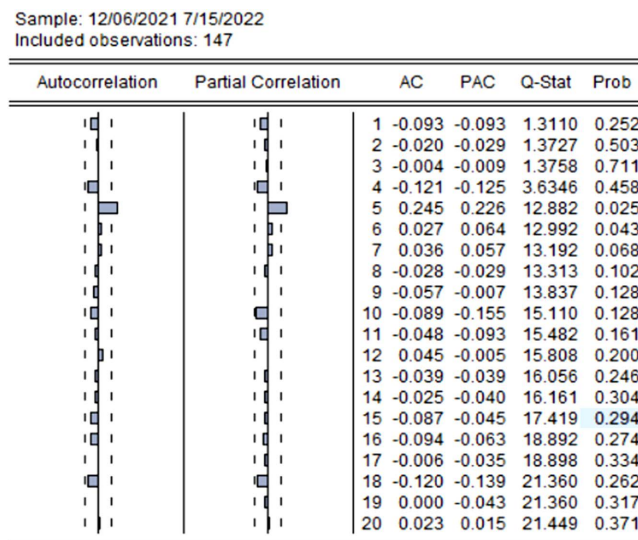


Figure 4. Differential post-sequence ACF and PACF plots  
图 4. 差分后序列 ACF 和 PACF 图

Table 4. AIC values corresponding to different orders  
表 4. 不同阶数对应的 AIC 值

$p$	$d$	$q$	AIC
2	1	2	11.361
1	1	1	11.789
1	1	0	9.968
5	1	5	0.016

根据上文，建立 ARIMA(5,1,5)模型进行拟合，结果如图 5 所示，模型对应的  $P$ 值 =  $0 < 0.05 = \alpha$ ，故模型显著有效。进行参数估计，并写出对应表达式为：

$$Y_t = -0.371419y_{t-3} - 0.335192y_{t-5} + 0.384267\varepsilon_{t-3} + 0.614337\varepsilon_{t-5} + \varepsilon_t. \quad (2)$$

### 3.4. 模型的显著性检验

模型的显著性检验主要检验模型的有效性，即拟合模型是否充分提取观测值序列中所有的样本相关信息。当拟合模型为有效模型时，残差序列  $\varepsilon_t$  不存在任何相关性，即残差序列  $\varepsilon_t$  为白噪声序列。对模型进行残差检验，由图 6， $P$ 值 =  $0 > 0.05 = \alpha$ ，故不能拒绝原假设，即残差为白噪声序列，模型显著有效。

此外，DW(Durbin\_Waston)检验也是检验残差的自相关方法[2]。检验统计量 DW 的取值范围为区间  $[0, 4]$ ，DW 趋近于 0，序列显著正相关；DW 趋近于 4，序列显著负相关；DW 趋近于 2，序列不存在自相关性。经检验得，DW = 2.13 趋近于 2，因此可判断残差为白噪声序列。

Sample (adjusted): 12/14/2021 7/15/2022  
 Included observations: 142 after adjustments  
 Convergence achieved after 30 iterations  
 MA Backcast: 12/07/2021 12/13/2021

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(3)	-0.371419	0.157242	-2.362084	0.0196
AR(5)	-0.335192	0.142787	-2.347491	0.0203
MA(3)	0.384267	0.121450	3.163998	0.0019
MA(5)	0.614337	0.127991	4.799834	0.0000

R-squared	0.104502	Mean dependent var	0.004296
Adjusted R-squared	0.085035	S.D. dependent var	0.251493
S.E. of regression	0.240563	Akaike info criterion	0.016094
Sum squared resid	7.986124	Schwarz criterion	0.099356
Log likelihood	2.857361	Hannan-Quinn criter.	0.049928
Durbin-Watson stat	2.136712		

Figure 5. ARIMA(5,1,5) fitting result

图 5. ARIMA(5,1,5)拟合结果

Sample: 12/14/2021 7/15/2022  
 Included observations: 142  
 Q-statistic probabilities adjusted for 4 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.073	-0.073	0.7713		
2	-0.046	-0.051	1.0769		
3	0.033	0.026	1.2419		
4	-0.114	-0.113	3.1736		
5	0.037	0.024	3.3764	0.066	
6	0.021	0.014	3.4444	0.179	
7	0.044	0.056	3.7331	0.292	
8	0.025	0.020	3.8254	0.430	
9	-0.018	-0.004	3.8765	0.567	
10	-0.020	-0.019	3.9357	0.685	
11	-0.052	-0.048	4.3585	0.738	
12	0.011	0.004	4.3791	0.821	
13	-0.056	-0.066	4.8685	0.846	
14	-0.031	-0.044	5.0235	0.890	
15	-0.102	-0.130	6.7002	0.823	
16	-0.062	-0.078	7.3314	0.835	
17	0.021	-0.012	7.4047	0.880	
18	-0.106	-0.113	9.2701	0.813	
19	-0.005	-0.042	9.2737	0.863	
20	0.048	0.028	9.6661	0.883	
21	-0.043	-0.022	9.9818	0.904	
22	-0.087	-0.108	11.269	0.883	
23	-0.012	-0.028	11.294	0.914	
24	0.040	0.032	11.574	0.930	

Figure 6. Residual series autocorrelation plot

图 6. 残差序列自相关图

观察图 7，模型的残差序列值基本拟合在直线  $y = 0$  附件呈对称分布，因此，也可判断残差序列不存在自相关。

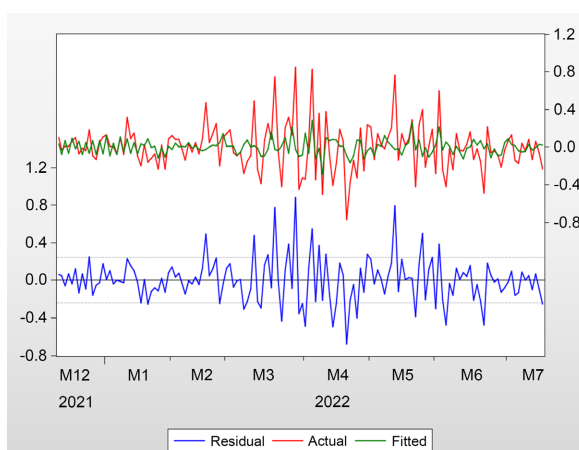


Figure 7. The final result

图 7. 最终效果图

### 3.5. 模型预测结果分析

由表 5 可知, 由上表可知, 预测误差率均控制在 5% 以内, 虽仍有些许误差, 但预测值的误差相对较小, 在可控范围内, 即认为该模型具有较好的拟合效果。

**Table 5.** Model prediction results

**表 5.** 模型预测结果

日期	实际值	预测值	误差
2022-07-18	7.43	7.327237	1.38%
2022-07-19	7.54	7.296828	3.22%
2022-07-20	7.58	7.305444	3.62%
2022-07-21	7.52	7.267384	3.35%
2022-07-22	7.49	7.199295	3.88%

## 4. 结语

股票价格预测一直备受瞩目, 本文采用 ARIMA 对所选取的数据样本进行了预测, 介绍了建模的基本方法和步骤, 通过平稳性, 白噪声等检验, 构建了 ARIMA 并进行了预测。结果表明, 该预测方法是可靠的, 对原序列信息提取充分, 拟合效果较好。但也看到预测值较真实值偏小, 且预测期数越长误差越大。因此相关从业者可选择此方法进行短期预测, 以为其相关投资决策作为参考。

## 参考文献

- [1] Zhang, G.S., Zhang, X.D. and Feng, H.Y.P. (2016) Forecasting Financial Time Series Using a Methodology Based on Autoregressive Integrated Moving Average and Taylor Expansion. *Expert Systems*, **33**, 501-516. <https://doi.org/10.1111/exsy.12164>
- [2] Su, W.B. (1997) The Supplement and Improvement of DW Test. *Application of Statistics & Management*, **2**, 37-41.
- [3] 刘琦. 基于 LS-SVM 的时间序列预测方法及其应用研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2011.
- [4] 潘雄锋, 彭晓雪. 时间序列分析[M]. 北京: 清华大学出版社, 2016.
- [5] 崔振辉, 李林川, 赵承利, 杨挺. 基于 ARIMA 的电力视频流量分析和预测[J]. 天津大学学报(自然科学与工程技术版), 2015, 48(1): 49-55.
- [6] 杨琦, 曹显兵. 基于 ARMA-GARCH 模型的股票价格分析与预测[J]. 数学的实践与认识, 2016, 46(6): 80-86.
- [7] 吴玉霞, 温欣. 基于 ARIMA 模型的短期股票价格预测[J]. 统计与决策, 2016(23): 83-86.
- [8] 蒋奇, 刘永文. 基于 ARIMA 模型的美元汇率预测[J]. 经济研究导刊, 2022(20): 69-71.
- [9] 黄荣庚, 龙静, 潘志刚, 陈焕新, 刘江岩, 刘佳慧, 李正飞. 基于 ARMA 模型的地铁站环控系统能耗预测[J]. 制冷学报, 2019, 40(1): 88-93.