

基于时间序列方法的股票价格分析

左鑫磊

广西师范大学数学与统计学院, 广西 桂林

收稿日期: 2023年12月19日; 录用日期: 2024年2月21日; 发布日期: 2024年2月27日

摘要

股票价格指数是一个国家经济建设健康状况的体温表, 它的变化大致反映了该国经济结构和经济活动的宏观变化趋势。因此, 选取合适的实证研究方法对股票价格指数进行预测分析, 具有重要的实证意义。近年来, 基于时间序列分析的预测方法在各个领域中都得到了广泛的应用。而对股票价格进行预测较为普遍的模型就是时间序列模型。所以本文以神州长城指数为例, 将时间序列建模方法应用于股票价格指数的建模与预测, 我们选择建立了ARIMA、GARCH模型进行拟合与预测。实验结果表明, 该模型的残差白噪声检验, 系数显著性检验都控制在一定范围内, 因此该模型拟合效果较好, 预测值接近实际值, 最后, 我们借助了该模型进行了股票指数未来一定时间内的预测。

关键词

时间序列分析, 股票价格预测, ARIMA模型

Stock Prices Based on Time Series Analysis

Xinlei Zuo

School of Mathematics and Statistics, Guangxi Normal University, Guilin Guangxi

Received: Dec. 19th, 2023; accepted: Feb. 21st, 2024; published: Feb. 27th, 2024

Abstract

The stock price index is a thermometer of the health of a country's economic construction. Its changes roughly reflect the macro trend of changes in the country's economic structure and economic activities. Therefore, it is of great empirical significance to select appropriate empirical research methods to conduct predictive analysis of stock price indexes. In recent years, forecasting methods based on time series analysis have been widely used in various fields. The most common model for predicting stock prices is the time series model. Therefore, this article takes the Sino-Great Wall Index as an example to apply the time series modeling method to the modeling and prediction of stock price index. We choose to establish ARIMA and GARCH models for fitting and

prediction. Experimental results show that the residual white noise test and coefficient significance test of this model are controlled within a certain range, so the model has a good fitting effect and the predicted value is close to the actual value. Finally, we used this model to conduct stock index forecasts within a certain period of time in the future.

Keywords

Time Series Analysis, Stock Price Prediction, 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. 绪论

1.1. 研究背景和意义

股票是一种由股份有限公司签发用以证明股东所持股份的凭证,它表明股票的持有者对股份有限公司的部分资本拥有所有权。由于股票具有一定的经济利益,且可以上市流通转让,因此股票也是一种有价证券。

1.1.1. 股票价格研究背景

股票价格研究是对证券市场运行机理的深入探索,有助于揭示价格形成的驱动因素和市场行为的规律。通过分析股票价格的波动、走势以及相关变量之间的关系,研究者可以发现价格形成的内在逻辑,从而为进一步理解市场提供了基础。例如,股票价格研究可以探究市场信息的有效性、投资者行为的决策规律以及市场风险的传导机制等,这对于深化资本市场理论以及监管政策的制定都具有重要意义。

1.1.2. 股票价格研究意义

股票价格研究在现实也有着广泛的应用价值。首先,它对于投资者的投资决策具有指导作用。投资者可以通过对股票价格的研究,了解股票的估值水平以及短期走势,为投资决策提供依据。股票价格研究包括技术分析和基本分析两种方法,技术分析主要关注价格走势的图表模式,而基本分析则从公司基本面入手,考察财务数据和行业环境等因素对股票价格的影响。通过合理运用这些研究方法,投资者可以更好地理解市场,选择适合自己风险偏好和投资目标的股票。其次,股票价格研究对于市场监管和风险控制也具有重要意义。证券市场的平稳运行需要监管部门能够及时发现市场异常、防范市场风险,而股票价格研究则可以提供监管机构和风险管理部门评估市场风险的依据。例如,基于对股票价格的研究,监管机构可以判断市场是否存在操纵行为、异常交易或者信息泄露等违法行为,及时采取措施维护市场秩序。此外,通过对市场涨跌幅、波动率等指标的研究,风险管理部门可以评估市场风险的程度,制定相应的风险管理措施,保护投资者的权益。

1.2. 国内外文献综述

时间序列分析是指对一系列依照时间顺序排列并连续观测得到的数据进行分析的统计方法。在实际生产和社会管理中,时间序列分析在经济、环境、医学、物流等多个领域都有着重要应用。时间序列模型方面,ARIMA模型仍然是时间序列分析中使用最广泛的模型之一。随着时间序列数据的不断增加,基于传统ARIMA [1]模型的拟合效果逐渐降低,为了提高模型的准确性,人们提出了各种改进的时间序列模型,如ARCH/GARCH模型。此外,VAR模型在多种时间序列变量之间建立统一模型,能够捕捉它们

之间的相互影响关系[2]。

多元时间序列分析方面，通过同时建立多个时间序列的联合模型，可以充分挖掘时间序列之间的相关性。传统的多元时间序列方法包括 ARIMAX [3]和 VAR [4] [5]等，近年来，越来越多的研究者使用基于深度学习的方法，如 LSTM [6]、GRU [7]等，来解决多元时间序列的建模问题。

深度学习方面，LSTM 是其中的代表之一。LSTM 不仅能够通过记忆功能建模长期依赖性，还具有记忆稳定性，在时间序列预测任务中表现出色。随着神经网络的不断发展，2018 年提出的 Transformer [8]网络也在时间序列领域得到了广泛应用，其在关键任务上表现优异，如股票预测、自然语言处理等。

应用领域方面，时间序列分析在很多领域涉及到决策制定，如金融风险管理、环境治理、医疗诊断等。近年来，随着物流、生产等领域的数字化转型，时间序列分析在预测订单量、库存管理等方面也得到广泛的应用。

总的来说，时间序列分析领域在方法和应用上都有不断的创新和发展。未来，随着大数据和深度学习技术的不断普及，时间序列分析将会更加有效地应用于实际生产和社会管理中的各个领域

2. 模型

本文在这一章主要介绍时间序列的相关概念以及模型，以便为一下个章节的顺利开展做一个铺垫。时间序列是指以时间顺序形态出现的一连串观测值集合，或更确切的说，对某动态系统随时间连续观察所产生有序的观测集合。

2.1. ARMA 模型

ARMA 模型的全称是自回归移动平均模型，它是目前最常用的拟合平稳序列的模型，它可以细分为 AR 模型、MA 模型和 ARMA 模型三大类。具有如下结构的模型称为自回归移动平均模型，简记为 ARMA(p, q)，

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \\ \phi_p \neq 0, \theta_q \neq 0, \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t, \\ E(x_s \varepsilon_t) = 0, \forall s < t. \end{cases}$$

特别的，当 $\phi_0 = 0$ ，称为中心化 ARMA(p, q)模型。引进延迟算子，ARMA(p, q)模型简记为， $\Phi(B)x_t = \Theta(B)\varepsilon_t$ 。式中 $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ，为 p 阶自回归系数多项式 $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ，为 q 阶移动平均系数多项式。

显然，当 $p = 0$ 时，ARMA(p, q)模型就退化成了 MA(q)模型，当 $q = 0$ 时，ARMA(p, q)模型就退化成了 AR(p)模型。所以，AR(p)模型和 MA(q)模型实际上就是 ARMA(p, q)的特例，它们都统称为 ARMA 模型。而 ARMA(p, q)模型的统计性质也正是 AR(p)模型和 MA(q)模型统计性质的结合。

2.2. ARCH 与 GARCH 模型

ARCH 模型的全称是自回归条件异方差模型，有时简称为条件异方差模型。该模型的构造原理如下：假设在历史数据已知的情况下，零均值、纯随机残差序列具有异方差性，其完整结构为：

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 \\ e_t \stackrel{i.i.d}{\sim} N(0, 1) \end{cases}$$

ARCH 模型的实质是使用误差平方序列的 q 阶移动平均拟合当期异方差函数值。由于移动平均模型具有自相关 q 阶截尾性，所以 ARCH 模型实际上只适用于异方差函数短期自相关过程。

在实践中，有些残差序列的异方差是具有长期自相关性的，这时如果用 ARCH 模型拟合异方差函数，将会产生很高的移动平均阶数，这会增加参数估计的难度并最终影响模型的拟合精度。为了修正这个问题，Bollerslov 在 1985 年提出了广义自回归条件异方差模型，它的结构如下：

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 + \sum_{i=1}^p \eta_i h_{t-i} \\ e_t \stackrel{i.i.d.}{\sim} N(0,1) \end{cases}$$

GARCH 模型实际上就是在 ARCH 的基础上，增加考虑了异方差函数的 p 阶自相关性而形成的；它可以有效地拟合具有长期记忆性的异方差函数。显然 ARCH 模型是 GARCH 模型的一个特例，ARCH(q) 模型实际上就是 $p=0$ 的 GARCH(p,q) 模型。

3. 基于时间序列的股票实例分析

3.1. 神州长城股票收盘价格建模

3.1.1. 数据预处理

图 2 是神州长城开盘价的自相关与偏自相关图，从中可以看出神州的开盘价序列其自相关函数衰减是非常缓慢的。从图 1 和图 2 可以判断神州长城的开盘价序列是一个非平稳时间序列。为了消除开盘价时间序列的非平稳性，可以对序列进行一阶差分。

从图 3 可以看出神州长城开盘价一阶差分序列基本上是围绕 0 值上下波动，初步判定其是一个平稳的时间序列，下面再通过自相关及偏自相关函数来进一步检验平稳性。

从图 4 可以判断，神州长城股价时间序列的一阶差分序列是一个平稳时间序列。通过一阶差分消除



Figure 1. Time series of opening price

图 1. 收盘价时间序列图

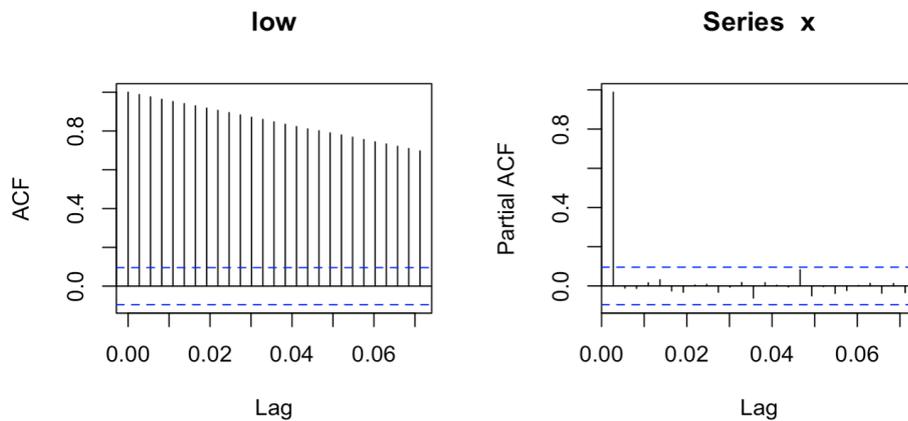


Figure 2. Autocorrelation and partial autocorrelation

图 2. 序列自相关图和偏自相关图

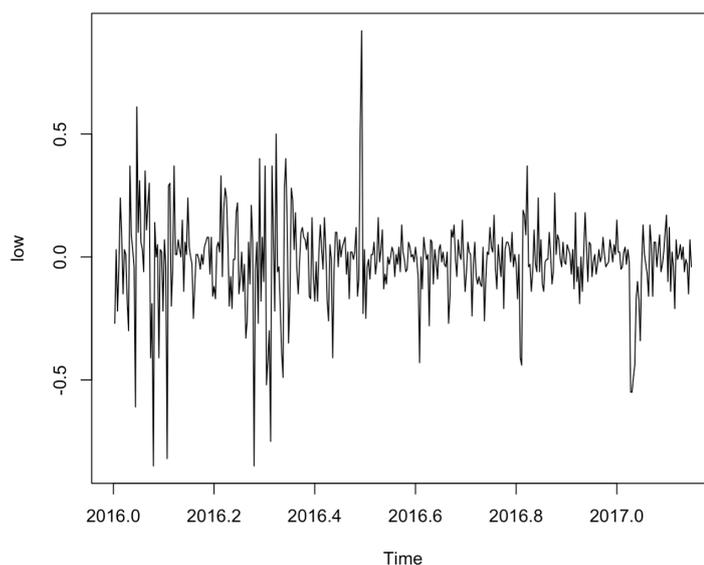


Figure 3. Close price: difference of first order

图 3. 收盘价一阶差分图

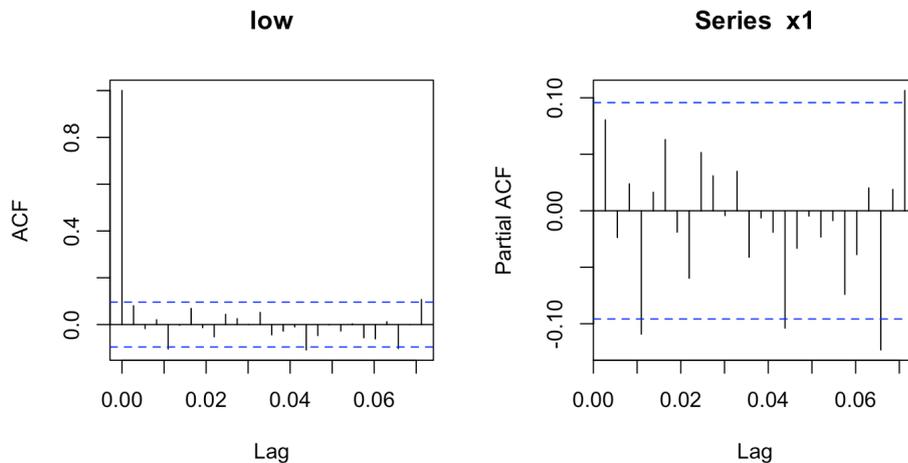


Figure 4. Autocorrelation and partial autocorrelation

图 4. 自相关图和偏自相关图

了线性趋势，而且由图看出序列的自相关性不明显，根据平稳性判断准则，原始非平稳数据经过一阶差分能迅速转化为平稳序列，说明原始序列的非平稳性较弱。

3.1.2. 建立 ARIMA 模型

通过一阶差分消除了线性趋势后，由一阶差分序列的自相关图可知其自相关函数是一阶截尾的，所以我们可以先对原始收盘价序列建立 ARIMA(0,1,1)模型，即对差分后的序列建立 MA(1)模型。

Table 1. White noise test)

表 1. 残差白噪声检验信息表

1. $i=1$	2. X-squared = 0.1105, df = 1, p-value = 0.7395
3. $i=2$	4. X-squared = 1.5031, df = 2, p-value = 0.4716
5. $i=3$	6. X-squared = 3.5689, df = 3, p-value = 0.3119
7. $i=4$	8. X-squared = 4.3636, df = 4, p-value = 0.359
9. $i=5$	10. X-squared = 4.7597, df = 5, p-value = 0.4459
11. $i=6$	12. X-squared = 5.1010, df = 6, p-value = 0.5309

表 1 为白噪声检验结果，由表中计算所得的 p 值可知该拟合模型的残差白噪声检验显示该模型显著成立，利用该拟合模型，还可以预测序列未来的水平，如图 5。

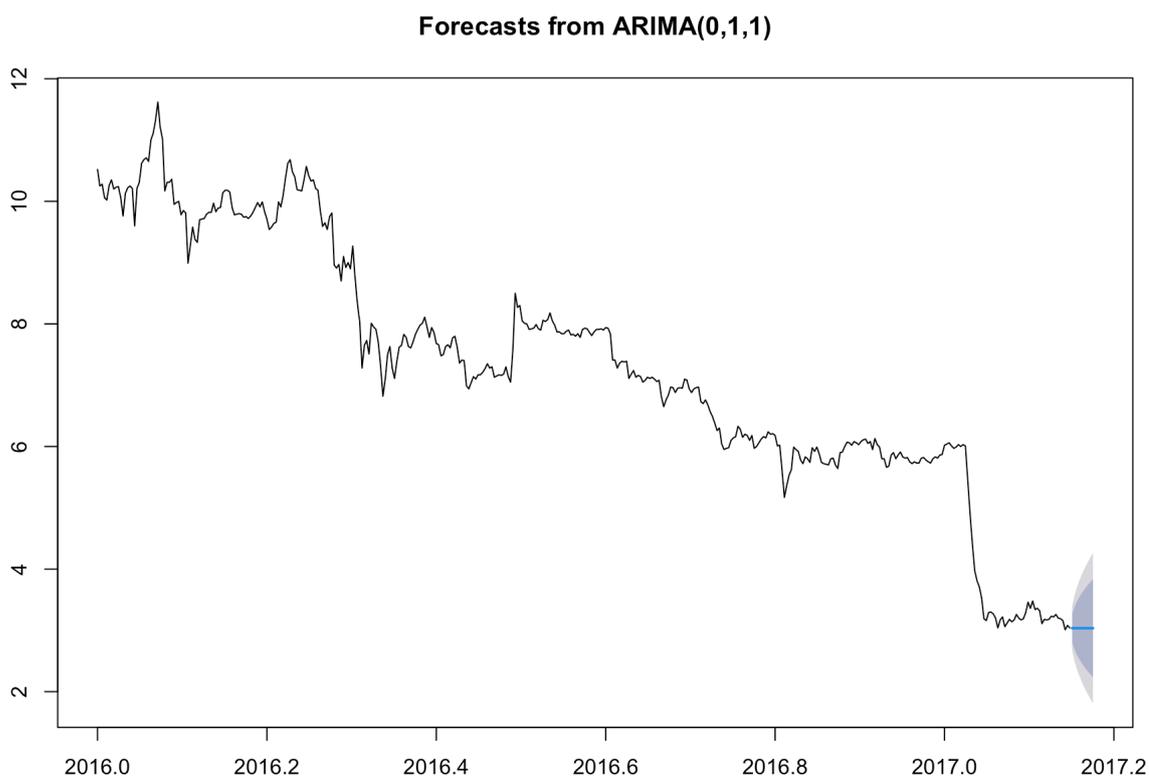


Figure 5. Time series prediction chart of the closing price

图 5. 神州长城收盘价的序列预测图

虽然使用 ARIMA 模型能拟合出非平稳序列，但是其对残差序列有一个重要假定——残差序列 $\{\epsilon_t\}$ 为零均值白噪声序列。我们用拟合的模型拟合出股票价格再计算出残差序列如图 6。

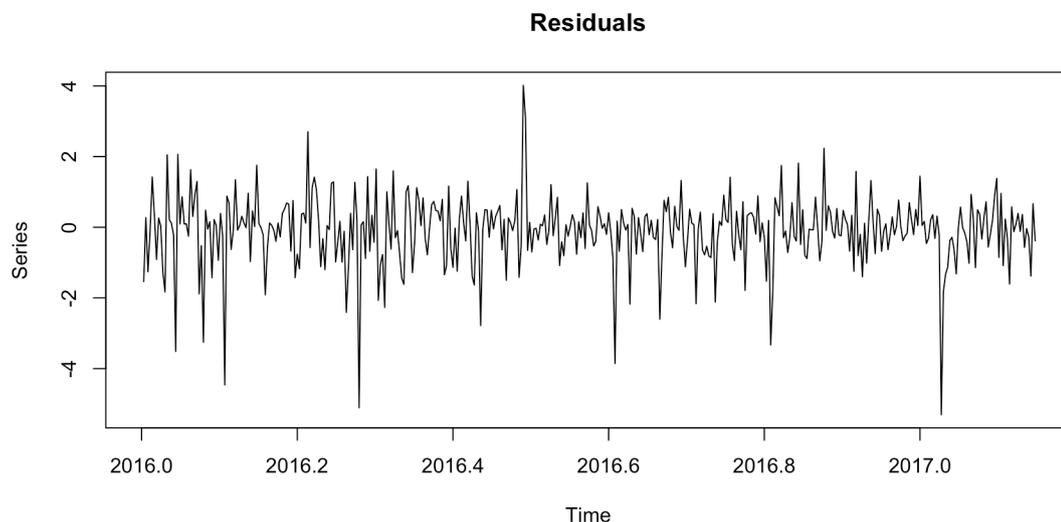


Figure 6. Residual time series of ARMA model
图 6. ARMA 模型的残差序列图

残差序列要满足如下三个假定条件：1) 零均值，如图 3~6 残差序列的值基本上是围绕 0 值上下波动的。2) 纯随机，如表 1 该拟合模型的残差白噪声检验 P 值均显著大于显著水平 α ，所以该序列不能拒绝原假设，即残差为白噪声序列。3) 方差齐性，如图 6 我们可以看出残差序列方差一直在波动，不再是个常数，它会随着时间变化而变化。所以我们认为我们建立的 ARIMA(0,1,1)模型存在异方差，接下来便使用 GARCH 模型提取异方差中蕴涵的相关信息。

3.1.3. 建立 GARCH(1,1)模型

根据 GARCH 模型建立步骤，初步确定收盘价时间序列的基本形式，并对残差进行拉格朗日乘子检验，检验时间序列是否存在 ARCH 效应：如果符合，根据相关效应及 GARCH 模型一般形式，确立几 GARCH 模型。根据上述分析，可尝试采用 AR-GARCH 模型对原始序列进行拟合，结合自相关图我们已经初步确定对神州长城收盘价时间序列建立 ARIMA(0,1,1)模型，如果 ARIMA(0,1,1)模型存在 ARCH 效应，则对其残差建立 GARCH(1,1)模型。

Table 2. Conditional heteroscedasticity test
表 2. 条件异方差检验结果信息

13. $i = 1$	14. X-squared = 38.427, df = 1, p-value = 5.683e-10
15. $i = 2$	16. X-squared = 49.091, df = 2, p-value = 2.188e-11
17. $i = 3$	18. X-squared = 57.761, df = 3, p-value = 1.768e-12
19. $i = 4$	20. X-squared = 66.689, df = 4, p-value = 1.134e-13
21. $i = 5$	22. X-squared = 69.546, df = 5, p-value = 1.273e-13
23. $i = 6$	24. X-squared = 69.900, df = 6, p-value = 4.288e-13

再对 ARIMA(0,1,1)进行 LM 检验得， $LM = \frac{(SST - SSE)/q}{SSE/(T - 2q - 1)} = 38.091 > \chi_{0.05}^2(1) = 3.841$ 和 $p\text{-value} =$

$6.751e-10$ 。由 LM 统计量所对应的 P 值小于显著性水平 0.05 时，拒绝原假设，认为该序列方差非齐性。所以波动信息的提取首先是考察 ARIMA(0,1,1)模型的残差平方序列的异方差特征。Portmanteau Q 检验也显示残差序列显著方差非齐，且具有长期相关性。所以构造 GARCH(1,1)模型，并根据该模型的拟合绘制

波动的 95%置信区间图和方差齐性置信区间比较图。

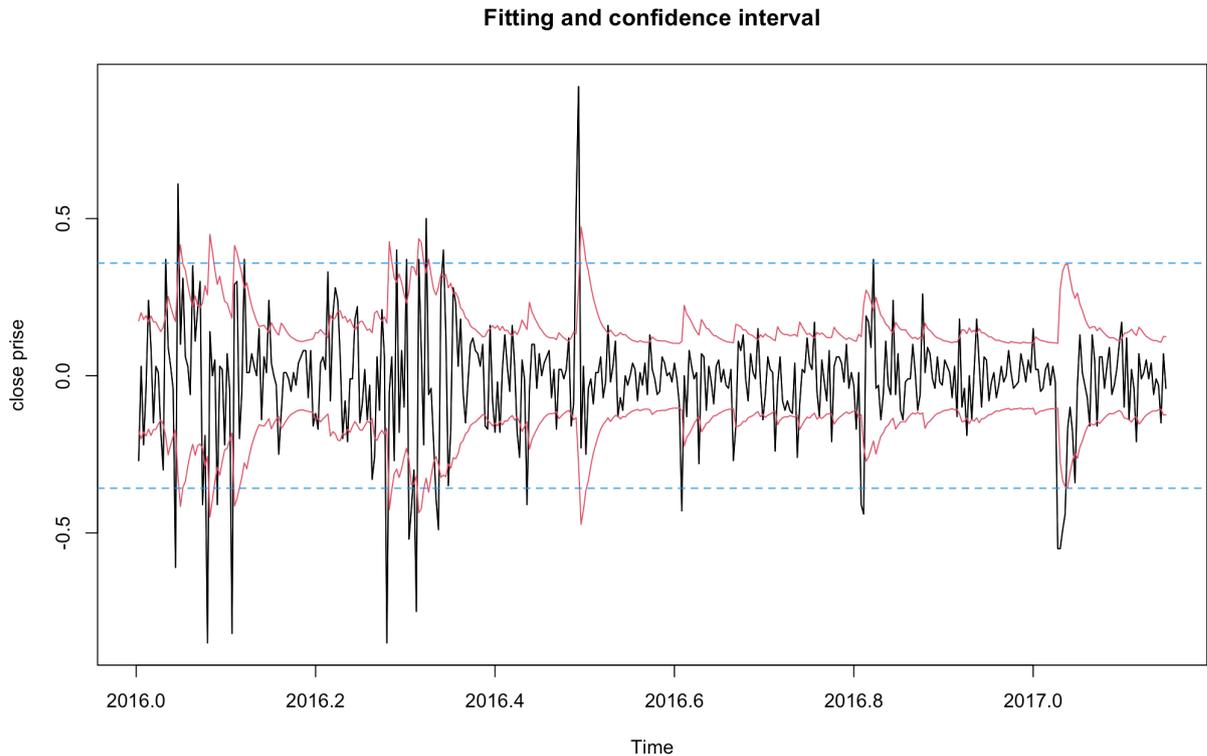


Figure 7. Fitting and confidence interval

图 7. 拟合与置信区间

图 7 中间的波动曲线是一阶差分时序图，两条平行虚线是根据方差齐性得到的 95%置信区间，两条波动实线为根据 GARCH(1,1)模型拟合结果绘制波动 95%的置信区间。综合水平模型和波动模型，我们得到的完整拟合模型为：

$$\begin{cases} \nabla x_t = \varepsilon_t + 0.1262\varepsilon_{t-1} \\ v_t = \sqrt{h_t}e_t \\ h_t = 0.7592h_{t-1} + 016231v_{t-1}^2 \end{cases}$$

在 GARCH 模型中参数 α ， β ，分别是前一期波动率的估计系数和估计残差的系数，单独看二者没多大的经济意义，主要看 $\alpha + \beta$ 这两个系数之和，系数越小，则说明该股票的波动相对平缓稳定一些，即 GARCH(1,1)模型中的 $\alpha + \beta$ 都小于 1，在模型满足平稳条件下，说明随机冲击对股市影响的持续时间是有限的，能够对其进行各项检验。而神州长城股票收盘价所建立的 GARCH(1,1)模型的 $\alpha + \beta$ 的数值很大，非常接近于 1，说明我国股票市场波动对冲击的反应是以一个相对较慢的速率衰减。随机冲击的影响还是具有相当程度的持续性。当证券收益率一旦受到冲击出现异常波动，则在短期内很难得以消除。

3.1.4. 模型拟合预测

再根据我们建立的 GARCH 模型拟合趋势并预测未来十天神州长城股票收盘价，ARIMA(0,1,1)模型包含了水平线性的趋势和 GARCH(1,1)蕴含了误差的波动率，由对残差建立的 GARCH 模型我们可以得到未来十天预测的波动率，如表 3。

Table 3. Volatility
表 3. 波动率

25. 0.1100	26. 0.1168	27. 0.1228	28. 0.1281	29. 0.1328	30. 0.1370	31. 0.1407	32. 0.1441	33. 0.1472	34. 0.1500
------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

由我们得到的完整拟合模型预测未来 10 天股票收盘价为表 4。

Table 4. The closing price forecast for the next 10 days
表 4. 未来 10 天收盘价格预测

35. 3.09455	36. 3.15598	37. 3.165685	38. 3.12155	39. 3.13211	40. 3.15121	41. 3.148454	42. 3.26541	43. 3.20155	44. 3.18121
-------------	-------------	--------------	-------------	-------------	-------------	--------------	-------------	-------------	-------------

图 8 中黑色的虚线为原始数据的时间序列，而红色的实线为根据我们建立的模型拟合的时间序列，可见拟合效果还是不错的。

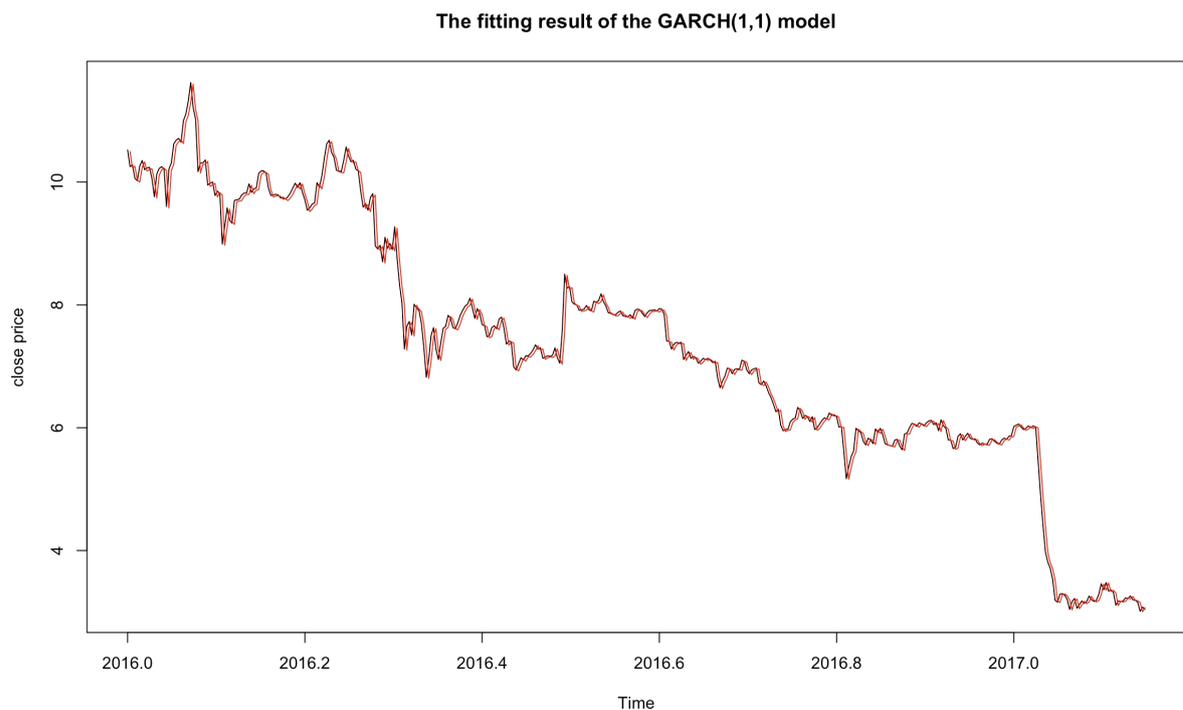


Figure 8. The fitting result of the GARCH (1,1) model
图 8. GARCH(1,1)模型拟合图

3.2. 股票日成交量建模

股票成交量为买卖双方达成交易的数量。成交量是市场供需的表现，是判断市场走向的重要指标；当市场行情持续上涨，出现急剧增加的成交量，持续上涨之后，股价在高位大幅度震荡，卖压沉重，成交量减少，从而形成股价下降的趋势；股价急剧下降之后，又会形成大量的成交量。股价一般会随着成交量的上涨而上升，因此，研究股票的成交量对股市行情也有着重要的作用。

3.2.1. 序列特征初分析

首先，画出成交量的时序图，如图 9 所示。该序列无明显趋势效应以及周期性变化，可近似看做平稳序列。进一步，对神州长城股票日成交量数据我们对其作纯随机性检验。根据假设检验，提出检验的原假设与备择假设： $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, \forall m \geq 1$ vs. $H_1: \text{至少存在某个 } \rho_{k \neq 0}, \forall m \geq 1, k \leq m$ 。

在 400 多个数据下，其近似为大样本数据，根据 Box 和 Pierce 提出的 Q 检验统计量，由观测数据我们得到了该序列延迟 12 期的样本自相关系数，如表 5。

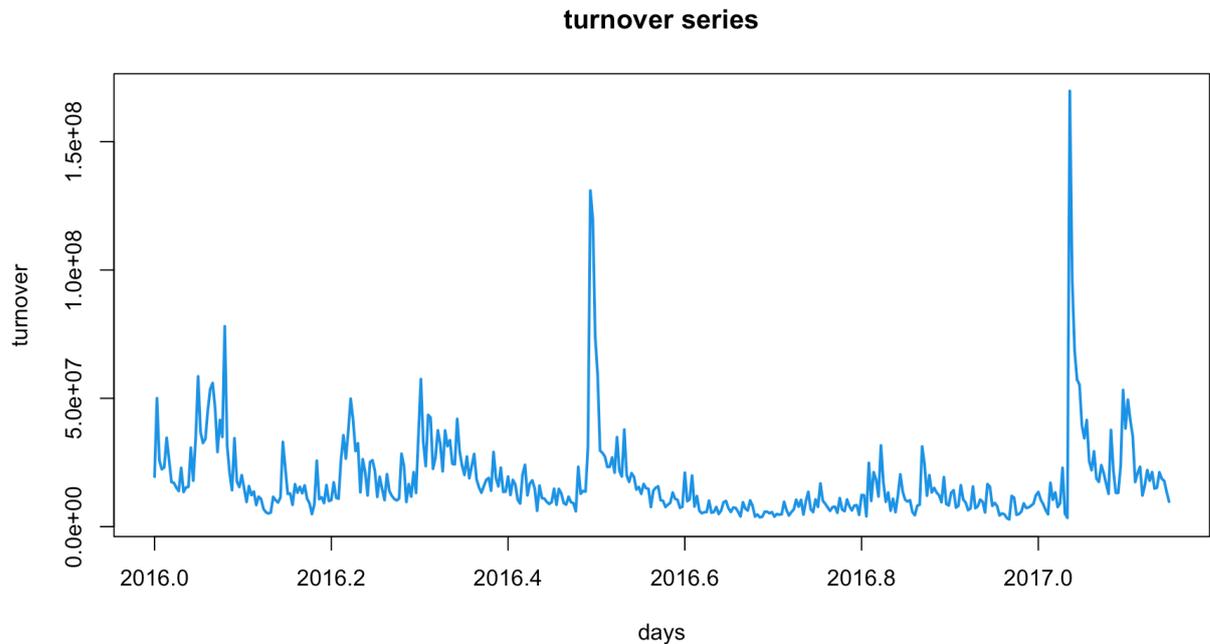


Figure 9. Time series chart of daily trading volume

图 9. 日成交量时序图

Table 5. Stock daily trading volume sequence delay 12 self-correlation coefficient

表 5. 股票日成交量序列延迟 12 自相关系数

延迟期数	1	2	3	4	5	6
ρ_k	0.999288	0.654965	0.476633	0.420983	0.346437	0.2812
延迟期数	7	8	9	10	11	12
ρ_k	0.241812	0.223217	0.18125	0.145774	0.140234	0.124476

根据上表，容易计算出表 6 的结果。表 6 中 Q 检验统计量结果显示，延迟期数为 6 期与 12 期时在显著性水平 α 下，统计量的值均大于 $\chi_{1-\alpha}^2(m)$ 分位点；则拒绝原假设，认为该序列为非白噪声序列。序列的白噪声检验通过之后，并进一步画出其自相关图与偏自相关图。

Table 6. Q-test result

表 6. Q 检验统计量结果

延迟	统计量检验	
	Q 统计量值	$\chi_{0.95}^2(m)$
延迟 6 期	853.04	12.5916
延迟 12 期	936.02	21.0261

根据序列的自相关图与偏自相关图显示，我们认为该序列自相关系数拖尾，偏自相关系数 3 步截尾；故可初步采用自回归 AR(3)模型来拟合该序列的发展趋势。

3.2.2. 建立 AR(3)模型

Table 7. Calculation results of AR (3) model
表 7. 模型相关计算结果汇总表

Coef	AR1	AR2	AR3	intercept
	1	2	3	4
s.e	0.999288	0.654965	0.476633	0.420983
σ^2		1.502e+14		
AIC		14,912.58		

根据计算 R 语言运算结果，我们得到表 7 结果，于是我们得到初步的模型为：

$$x_t - 18021096 = \frac{\varepsilon_t}{1 - 0.593B + 0.0119B^2 - 0.1443B^3}, \quad \varepsilon_t \sim N(0, 1.502 \times 10^{14})$$

或者等价写为：

$$x_t = 4939582 + 0.593x_{t-1} - 0.0119x_{t-2} + 0.1443x_{t-3} + \varepsilon_t \sim N(0, 1.502 \times 10^{14}),$$

延迟 6 期和延迟 12 期的 p 值约等于 1，即模型的残差白噪声检验 p 值接近 1，于是接受原假设，认为该拟合序列的残差为纯随机学咧；系数的显著性检验 1、3 均通过，第二个系数没有通过，模型有待优化。

3.2.3. 模型拟合及预测

根据上述模型，我们进行与原模型的拟合与 18 期预测，预测结果如表 8。

Table 8. 18 period of forecast results
表 8. 18 期预测结果

预测期数	1	2	3	4	5	6
预测值	17,876,398	14,750,108	11,996,797	17,916,059	15,646,686	13,648,057
预测期数	7	8	9	10	11	12
预测值	17,944,849	16,297,511	14,846,707	17,965,748	16,769,945	15,716,806
预测期数	13	14	15	16	17	18
预测值	17,980,919	17,112,885	16,348,412	17,991,931	17,361,825	16,806,894

为了能够给出预测值 95% 的置信区间，我们须计算出相应的 Green 函数值，由 Green 函数的递推关系式：

$$\begin{cases} G_0 = 1 \\ G_j = \sum_{k=1}^j \phi'_k G_{j-k}, \quad \phi'_k = \begin{cases} \phi_k, & k \leq p \\ 0, & k > p \end{cases} \end{cases}$$

得到递推关系：

$$\begin{cases} G_0 = 1 \\ G_1 = \phi_1 G_0 = 0.5935 \\ G_2 = \phi_1 G_1 + \phi_2 G_0 = 0.3397 \\ G_j = \phi_1 G_{j-1} + \phi_2 G_{j-2} + \phi_3 G_{j-3}, \quad j \geq 3 \end{cases}$$

根据上述 $j \geq 3$ 时的递推关系，我们便能计算得到相应的 Green 函数值，再根据 l 步预测的方差公式

$Var[e_t(l)] = (1 + G_1^2 + \dots + G_{l-1}^2) \sigma_e^2$ 得到 l 步预测股票日成交量的 95% 置信区间为:

$$\left(\hat{x}_t(l) - 1.96\sqrt{Var[e_t(l)]}, \hat{x}_t(l) + 1.96\sqrt{Var[e_t(l)]} \right).$$

由上述分析, 我们将序列后半部分的真实值、模型拟合值、模型预测值、拟合值的 95% 置信上下限以及预测值的 95% 置信上下限绘制于图 10 中(中间黑色曲线为模型观察值序列, 中间绿色线条为模型拟合曲线, 蓝色散点为 95% 的置信上下限, 红色散点为预测值 95% 的上下限, 蓝色曲线为预测值曲线)。

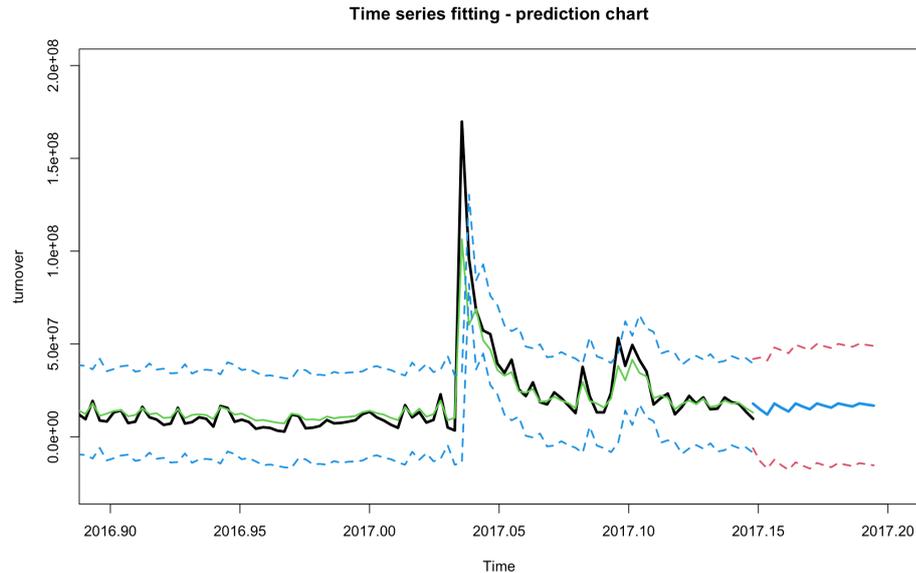


Figure 10. Time series fitting and prediction

图 10. 序列拟合预测图

3.2.4. 模型优化

上述模型系数的显著性检验并没有全部通过, 且预测值不太理想, 模型有待优化, 故进一步考虑对其进行差分处理, 再考察其差分序列的平稳性。经一阶差分后, 得到相应的时序图如图 11, 一阶差分后的序列较原序列显然更加平稳。

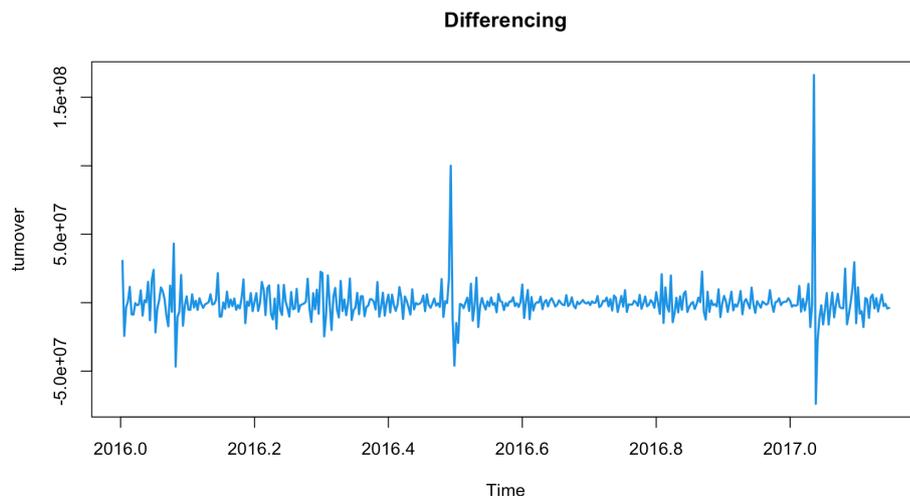


Figure 11. Difference of first order

图 11. 一阶差分图

再根据差分后的自相关图和偏自相关图定阶，自相关系数呈 2 阶截尾，偏自相关呈拖尾，故可采用 2 阶移动平均 MA(2)模型来拟合序列的发展趋势，相关结果如表 9，且相比于上一个模型，系数显著性检验以及残差白噪声检验在置信水平为 $\alpha = 0.05$ 的情况下，均通过了检验，从而得到模型，

$$\nabla x_t = 26799.67 + \varepsilon_t - 0.3726\varepsilon_{t-1} - 0.2731\varepsilon_{t-2}, \varepsilon_t \sim N(0, 1.582 \times 10^{14}).$$

Table 9. MA (1) model information
表 9. 模型相关信息表

	MA1	MA2	intercept
Coef	-0.3726	-0.2731	-26,799.67
s.e	0.0481	0.0549	219,159.90
σ^2		1.582e+14	
AIC		14,896.71	

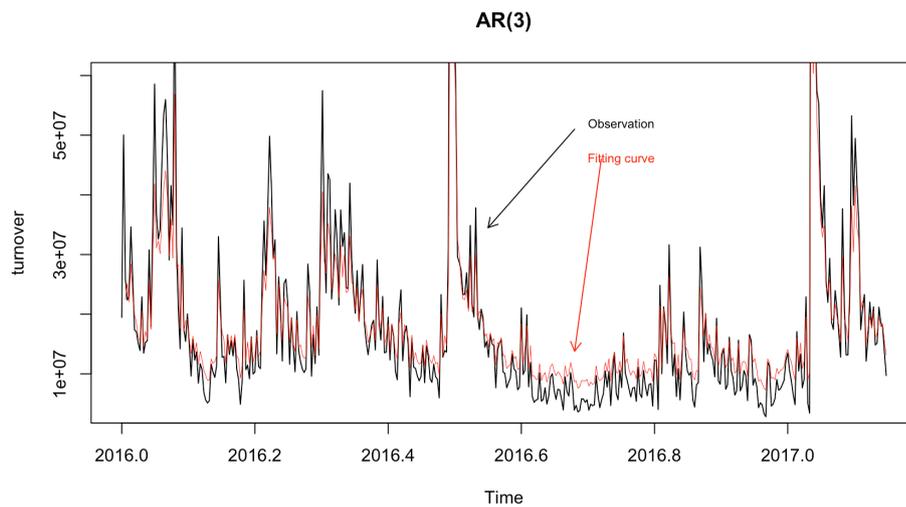
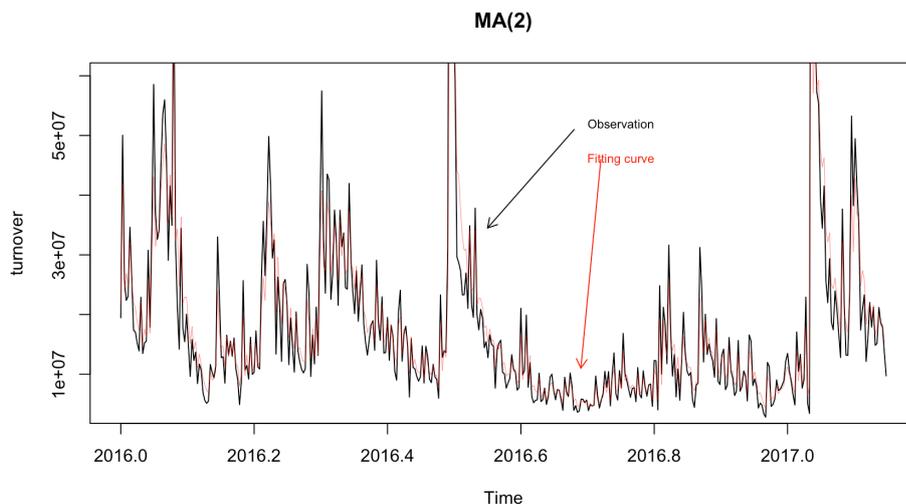


Figure 12. The fitting of AR (3) model and original sequence
图 12. AR(3)模型与原序列的拟合图



Fitting 13. Diagram of the MA (2) model after first-order difference and the original sequence
图 13. 一阶差分后 MA(2)模型与原序列的拟合图

为了能够对比两个模型的优劣性，我们分别将两个模型与原序列拟合，观察两个模型拟合的效果，从拟合的效果对比来看如图 12、图 13，其中不同颜色的曲线分别代表观测值曲线和拟合曲线。显然 MA(2)模型要优于 AR(3)模型，于是我们得到了较优的最终模型为 MA(2)模型。并利用此模型，预测未来五天的日交易量分别为：13,141,446、14,767,273、14,640,473、12,535,683、17,523,871。同时与真实未来五天的日交易量对比，我们预测值的平均误差仅在 35,213.91 左右，说明该模型不仅能够较好的拟合日交易量，并且预测也有较好的准确性。

4. 结论与启示

由于股价一般会随着成交量的变化而变化，因此本文对成交量进行分析，股票成交量的对象是神州长城的 400 多个股票成交量数据。首先我们先对数据进行纯随机性检验，判断其数据的平稳性。在 400 多个数据下，其近似为大样本数据，根据 Box 和 Pierce 提出的 Q 检验统计量，由观测数据，我们得到了该序列延迟 12 期的样本自相关系数，根据检验统计量的结果显示，可以认为该序列为非白噪声序列。序列的白噪声检验通过之后，我们根据该序列的时序图可以看出该序列无明显的趋势效应以及周期性变化，可近似看做平稳序列。接着根据序列的自相关图与偏自相关图，我们认为该序列自相关系数拖尾，偏自相关系数 3 步截尾；故可初步采用自回归 AR(3)模型来拟合该序列的发展趋势。在模型的残差白噪声检验和系数的显著性检验中，模型的残差白噪声检验 P 值均接近于 1，于是接受原假设，认为拟合序列的残差为纯随机序列；系数的显著性检验 1、3 均通过，第二个系数没有通过，模型有待优化。根据上述模型，我们还是可以进行对原模型的拟合与 18 期预测，计算出相应的 Green 函数值，于是得到步预测神州长城日成交量的 95%置信区间。

由于上述模型系数的显著性检验并没有全部通过，我们的模型有待优化，故后续还可以进一步考虑对其进行差分处理，再考察其差分序列的平稳性。通过一阶差分后的序列较原序列显然更加平稳，其自相关系数呈现 2 阶截尾，偏自相关图拖尾，故可采用 2 阶移动平均 MA(2)模型来拟合序列的发展趋势。而该模型也通过了残差白噪声检验以及系数的显著性检验。

随着机器学习的发展，神经网络在时间序列模型中的应用非常广泛，可以用于预测、分类和生成时间序列数据。RNN 是最常用的时间序列模型之一，它被设计用于处理具有时间依赖性的数据。长短期记忆网络(LSTM)：LSTM 是 RNN 的一个变体，专门用于解决长时间依赖问题。LSTM 在 RNN 的基础上增加了门控机制，能够更好地捕捉和记忆长期的依赖关系，这使得 LSTM 在处理长时间序列任务上表现出色。还有其他各种类型的神经网络可以根据实际情况选择使用，这也是我们后续的研究方向。

参考文献

- [1] 王振龙, 胡永宏. 应用时间序列分析[M]. 北京: 科学出版社, 2007.
- [2] 沃尔特, 恩德斯. 应用计量经济学: 时间序列分析[M]. 北京: 高等教育出版社, 2006.
- [3] 汪远征, 徐雅静. 多元平稳时间序列 ARIMAX 模型的应用[J]. 统计与决策, 2007(18): 132-135.
- [4] Granger, C.W.J. (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica: Journal of the Econometric Society*, **37**, 424-438. <https://doi.org/10.2307/1912791>
- [5] Johnson, R.A. and Wichern, D.W. (2002) *Applied Multivariate Statistical Analysis*. Pearson, London.
- [6] Shi, X., Chen, Z., Wang, H., et al. (2015) Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R., Eds., *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, Montreal, 7-12 December 2015, 802-810.
- [7] Chung, J., Gulcehre, C., Cho, K.H., et al. (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv:1412.3555.

- [8] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, 4-9 December 2017, 6000-6010.