

基于时间序列算法的资金流入流出预测模型比较分析

汪恺旻, 张 燕

北京师范大学珠海分校, 广东 珠海

收稿日期: 2022年4月24日; 录用日期: 2022年5月5日; 发布日期: 2022年5月27日

摘 要

商业公司的金融服务平台通常有着上亿的客户, 庞大的用户群体使得其服务场景中每天都必然会触及大量资金的流入和流出, 对资金的管控压力也越来越大。在当前的宏观经济环境和现实状况下, 要确保资本管理在一个可控的范围内, 对未来的资本流量规模做出预估十分必要。现有文献表明, 在实际操作中有多种模型可以进行相关的预测。使用比较多的包括基于统计学方法来完成的ARIMA模型、周期因子时间序列预测模型以及基于神经网络的LSTM时间序列预测模型等。本文以余额宝平台用户资金流量为背景, 通过比较以上三种不同资金流量预测模型的适用条件、优点及局限性, 尝试对比分析出针对不同条件下更加准确有效的资金流入流出预测模型, 最大程度贴近未来资金流量真实值, 从而更好地进行企业的经营管理。

关键词

资金流预测, ARIMA, 周期因子, 时间序列模型

Comparative Analysis of Capital Inflow and Outflow Prediction Models Based on Time Series Algorithm

Kaimin Wang, Yan Zhang

Beijing Normal University (Zhuhai), Zhuhai Guangdong

Received: Apr. 24th, 2022; accepted: May 5th, 2022; published: May 27th, 2022

Abstract

The financial service platform of commercial companies usually has hundreds of millions of

customers. The huge user group makes it inevitable to touch a large amount of capital inflow and outflow every day in its service scenario, and the pressure on capital control is also increasing. In the current macroeconomic environment and current situation, it is very necessary to predict the scale of future capital flow in order to ensure that capital management is within a controllable range. The existing literature shows that there are many models that can make relevant predictions in practical operation. ARIMA model based on statistical method, periodic factor time series prediction model and LSTM time series prediction model based on neural network are widely used. Taking the user capital flow of Yu'e Bao platform as the background, by comparing the applicable conditions, advantages and limitations of the above three different capital flow prediction models, this paper attempts to compare and analyze a more accurate and effective capital inflow and outflow prediction model under different conditions, which is close to the real value of future capital flow to the greatest extent, so as to better carry out the operation and management of enterprises.

Keywords

Capital Flow Prediction, ARIMA, Periodic Factor, Time Series Model

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

在金融市场的业务场景中每天都涉及大量的资金流入和流出, 面对庞大的用户群, 资金管理压力会非常大。在既保证资金流动性风险最小, 又满足日常业务运转的情况下, 精准地预测资金流入流出情况变得尤为重要。

支付宝所推出的余额宝作为一个货币的理财工具, 可以随借随取, 使得用户能够更加方便地管理自己的资金。每天大量的资金流入和流出对于资金管理的压力很大, 因此需要进行一个预测, 来判断未来一段时间资金流入流出的情况。如果预测准确, 对于用户的资金管理来说是非常有利的, 而如果预测结果偏差, 资金流的突然增加和减少都有可能造成一定的系统性风险, 而造成严重的损失。但是金融数据往往受经济、政治等多方因素的影响, 数据不稳定, 这给资金流量的预测带来了难度。

本文通过采用不同时间序列算法对余额宝用户的申购赎回数据的把握, 比较不同模型对于未来每日的资金流入流出情况的预测精准性, 从而分析不同算法模型在对应场景中的适应性与泛化性。对于货币基金而言, 资金流入意味着用户申购行为, 资金流出意味着用户赎回行为。

2. 文献综述

金融时间序列预测研究是根据过去若干个历史数据的波动, 建立模型来描述金融结构并预测时间序列上未来某一点的值。根据实际场景可以应用在不同的金融领域中, 比如对于股票、财政、税收等领域的资金预测, 这对于投资者和金融机构来说具有重要的研究意义, 除此之外也可以对电商流量或者各类理财产品利率进行预测[1]。

股票市场在金融领域占据重要的地位, 因此对股票和指数的预测一直以来都是海内外学者研究和探索的课题。但股票市场的价格走势收到经济、政治、军事等诸多因素的影响, 因此想要单纯的依靠理论分析就对于市场有精准的把控变得较为困难, 往往预测效果不佳。因此过多的研究倾向于使用统计学或

者人工智能的技术, 例如郝继升等基于自适应遗传算法优化的 BP 神经网络对股票价格进行了预测, 结果表明该模型在股票预测的精度方面得到了明显的提升[2]。Devi, B 等人的方法是使用 ARIMA 模型对股票趋势的有效时间序列进行分析预测, 使用了不同参数的 ARIMA 模型对于公司过去 5 年的股票数据进行了建模, 通过训练之后的时间序列模型来预测未来公司股价的波动趋势[3]。Adebiyi A A 的团队同样运用了 ARIMA 模型建立了股票预测模型, 通过对纽交所和尼日利亚证券交易所的股票数据进行收集和训练, 结果表明, 自回归移动平均模型在股票短期预测上有着一定的优势[4]。但同时 ARIMA 模型在使用上也有着一定的局限性, ARIMA 模型要求输入的时间序列是稳定的序列, 但在实际情况中大部分的金融数据并不一定是平稳的, 因此对于一些股价波动剧烈的股票类型来说, ARMA/ARIMA 模型并不具有泛用性和普适性。

金融时序的预测研究对各行业的资金流量预测应用也较为广泛。国内的研究如葛娜等基于 ARIMA 模型对某品牌鞋类的销售量进行预测, 分析结果表明该模型可以达到较好的效果[5]。陶佳等基于改进支持向量机的方法对于林业资金投资进行了预测, 与传统方法相比预测精度的提升显著[6]。赖慧慧使用 Prophet 模型对乘用车的消费税进行了预测, 这在行业分品目的税收预测上是一个全新的思路, 结果证明 Prophet 在预测具有季节性和节假日特征的税收预测的方面精度高, 具有优势[7]。黄莺等使用多维灰色模型 GM (1, N)模型和 Prophet 模型结合的方式对于电商行业的销售量进行了预测, 分析出这样的组合模型适用性更高[8]。国外研究例如 Safari 等将指数平滑模型和神经网络进行融合对石油价格进行预测研究, 结果表明相对于传统时间序列方法有较高的准确率[9]。

3. 模型概述

3.1. ARMA/ARIMA 模型

ARMA/ARIMA 时间序列模型是一组依赖于时间的随机变量。

1) ARMA 模型: (Auto Regressive Moving Average 自回归滑动平均模型)

ARMA 模型, 即 AR (Auto Regressive 自回归模型)和(MA Moving Average 滑动平均模型)模型的结合。相比于 AR 和 MA 模型, 它的准确度更高, 其存在 p, q 两个阶数, 称为 ARMA (p, q)模型, 表达式为:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + u_t - \varnothing_1 u_{t-1} + \varnothing_2 u_{t-2} - \dots - \varnothing_q u_{t-q}$$

其中, \varnothing 为自回归系数, u_t 表示白噪声, 是时间序列数值的随机波动。

ARMA 模型结合了 AR 和 MA 模型的特点, AR 解决了前后数据之间的关系, MA 则解决了随机变动即噪声的问题, 因此其可以拟合所有的时间序列。但是, 该模型是基于平稳数据建立的。

2) ARIMA 模型: (Auto Regressive, Integrated Moving Average 差分自回归滑动平均模型)

由于实际中很多数据都是不平稳的数据, 差分的过程就是把输入的不平稳数据平稳化, 即将前后两个数据进行减法操作。相较于 ARMA, ARIMA 增添了一个差分的操作, 将不平稳的数据进行差分平稳。ARIMA 具有三个阶数, p, d, q , 称为 ARIMA (p, d, q)模型, d 表示差分阶数。因此在对 ARIMA 模型进行拟合之前, 需要判断数据是否为平稳序列[10]。对于非平稳的数据需要先进行 d 阶差分运算, 化为平稳的时间序列。在使用 ARIMA 模型进行预测后还需要对预测数据进行还原的操作, 图 1 为该时间序列模型具体流程图。

3.2. Prophet 模型

Prophet 作为 Facebook 的一个开源的时间序列预测模型, 相比 ARMA/ARIMA 模型可以设置饱和和增长, 突变点, 节日与大事件等参数。在 Prophet 的模型预测中, 它不会给出一个确切的数字, 而会给出一个上下波动的范围, 这样的预测区间会更加合理, 并且不容易发生过拟合。

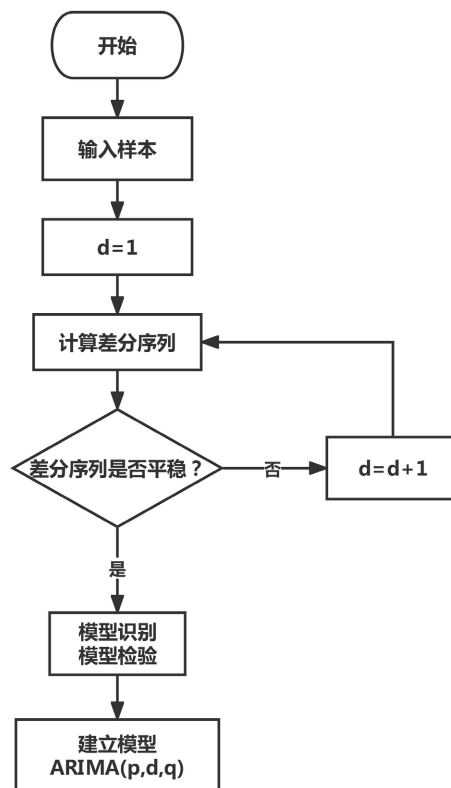


Figure 1. Specific flow chart of time series model

图 1. 时间序列模型具体流程图

Prophet 是一个基于加法模型(Additive Model)的时间预测, 其加法原理就是把一个原始数据拆成几个部分做累加的过程, 可以比较精准的去预测非线性的周期趋势。同时 Prophet 的优势在于添加了 Holiday 影响因子, 可以很好的对例如十一假期、春节等节假日带来的活跃数据的突变进行预测, 如中国传统节日春节, 在这个期间人们会购买大量新商品, 造成资金流发生明显的变化, 因此这些大事件前后的日期将会被单独考虑, 并且通过拟合附加的参数模拟节假日和事件的效果。

Prophet 的数学模型可以表示为:

$$y(t) = g(t) + s(t) + h(t) + e$$

其中, $g(t)$ 代表趋势项, 用来表示时间序列中非周期性的变化; $s(t)$ 代表周期项, 用来表示时间序列中周期性的变化。

3.3. 周期因子时间序列模型

很多时间序列都具有明显的周期性, 例如购物、客流量等, 对于通勤人数来说, 周一的流量最大而周末的流量最小, 类似于此类数据都会呈现出一个周期性的波动趋势, 因此需要考虑周期因素, 在预测之前需要确定周期长度, 比如一周或者一个月。

对于周期波动非常明显的的数据, 使用 ARMA/ARIMA 模型进行预测的准确度就会下降, ARIMA 的原理是用过去的几个时间点预测下一个时间点, 含有趋势性, 但运用在含周期因素的数据中, 预测结果就会出现较大的偏差。

因此使用周期因子预测时间序列的核心任务是尽可能准确的提取出周期特征 weekday, 具体操作为

获取周一到周日的均值, 再除以整体均值即可得到因子数。第二步需要设定一个基数 $base$, 可以选取最后一周的均值作为基数, 最后使用 $base$ 值乘上周期因子进行预测, 具体流程如图 2 所示:



Figure 2. Flow chart of periodic factor prediction
图 2. 周期因子预测流程图

对于不含趋势性相对稳定且周期性强的时间序列而言, 采取周期因子的时间序列预测会更合理。

4. 互联网金融资金流数据分析

本次研究数据来源于由阿里云提供的“资金流输入输出预测”大赛数据。数据集一共包含四个部分: 用户基本信息数据、用户申购赎回数据、收益率表和银行间拆借利率表。本文主要使用的是用户申购赎回数据表, 记录了 2.8 万用户在 427 天内的操作情况, 其中操作记录包括申购和赎回两个部分, 金额的单位为分, 即 0.01 元。数据经过了脱敏确保了安全性, 同时保证了: 今日余额 = 昨日余额 + 今日申购 - 今日赎回, 不会出现负值。

本文使用的数据集包含了余额宝资金流的详细信息。由于单个余额宝的交易记录是以用户维度来进行记录的, 但是单用户的余额宝使用行为缺乏规律性, 且各个用户的理财习惯和消费风格均有差异, 因此对于单个用户的资金流预测较为困难。基于此, 本文按照日期这一维度进行了预测和分析, 即对所有用户的交易记录进行了聚合和加总的操作, 预测分析的是所有用户每一天的总购买量与总赎回量[11]。对数据集进行聚合操作, 按照日期进行聚合, 将同一时间进行加和处理, 统计每一天的总的购买量与申购量, 得到一个 2×247 维的数据集。

分别提取申购量与赎回量的总体数据并将统计好后的数据可视化如图 3 所示:

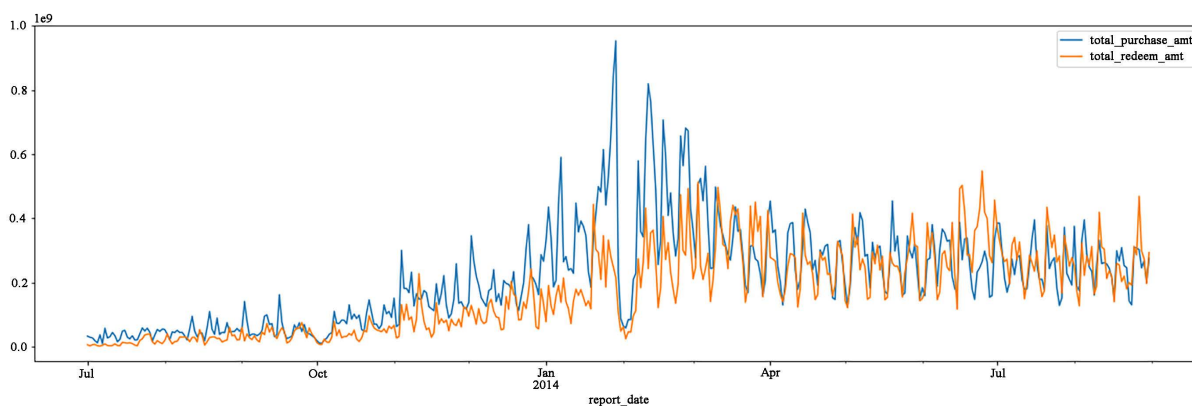


Figure 3. Time series of total daily purchases and redeems
图 3. 每日总购买与赎回量的时间序列图

在分析过程中, 为了明确模型的拟合效果, 并与真实数据进行对比计算误差, 需要将数据集分为训练集(train-data)和测试集(test-data), 本文选取前 13 个月即 2013 年 7 月 1 日到 2014 年 7 月 30 日的数据为训练集, 选取最后一个月即 2014 年 8 月份的数据为测试集进行模型的拟合。按照统计学的方法将时间序列进行拆解后, 对于指定时间范围的数据进行可视化分析, 分析结果如图 4 所示:

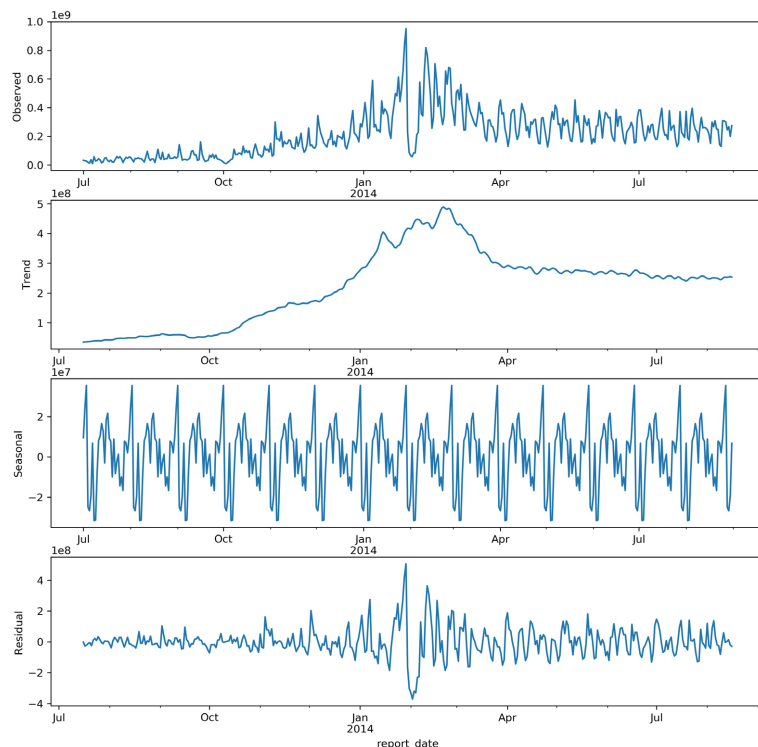


Figure 4. STL decomposition
图 4. STL 分解

由上图可以看出, 趋势性在一开始一直向上, 达到一定高峰后开始下降, 一开始的趋势很明显, 分析可得前期属于一个推广期, 用户数量不多, 而在后期随着用户数量的增加, 已逐渐趋于稳定。

5. 三个模型的预测比较

在使用时间序列模型 ARMA/ARIMA 建模之前需要对数据进行平稳性检测, 本文使用的检测方法为统计学中的 ADF 检测, 按照小概率事件原则进行检验。

ADF 检验需要进行假设检验, 假设检验过程如下所示:

原假设 H_0 : 序列不平稳; 备择假设 H_1 : 序列平稳。

如果统计所得的显著性检验统计量小于三个置信度 1%, 5%, 10%, 则分别对应有 99%, 95%, 90% 的把握拒绝原假设。分别对于申购量与赎回量数据集进行统计量计算可得, 原始数据的统计量分别为 -1.590 和 -1.372 , 大于置信度为 10% 的统计数值, 落入接受域, 因此没有充分理由拒绝原假设, 即证明原始数据是非平稳的。

对数据集进行一阶差分后再进行假设检验, 在进行差分后会出现空值, 因此还需要删除缺失值后再进行检验, 差分前后统计数值如表 1 所示:

Table 1. Numerical statistics before and after difference

表 1. 差分前后数值统计

	申购量 (purchase)	赎回量 (redeem)	统计结果
差分前(T 值)	-1.590	-1.372	非平稳
差分后(T 值)	-7.947	-11.692	平稳

由上述 ADF 平稳性检测可得, 在使用时间序列模型 ARIMA 进行预测时, 设置参数 d 为 1, 即需要进行一阶差分, 保证数据为平稳序列后在进行拟合[12]。

模型通过上述检验后, 使用训练集对模型进行拟合和预测, 为了更直观的说明预测精度, 将预测结果和测试集的实际结果进行对比如图 5 和图 6 所示来分析误差。

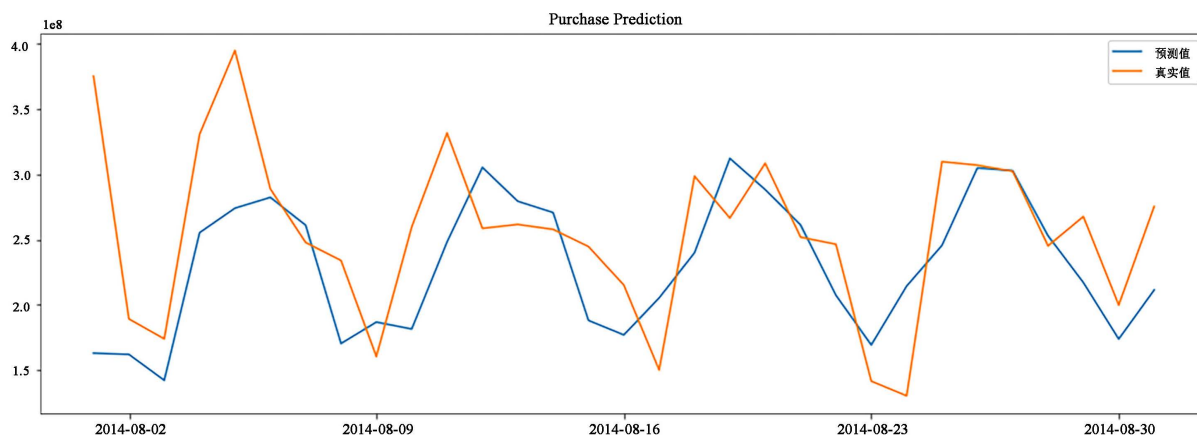


Figure 5. Prediction comparison of purchase model

图 5. 申购模型预测对比

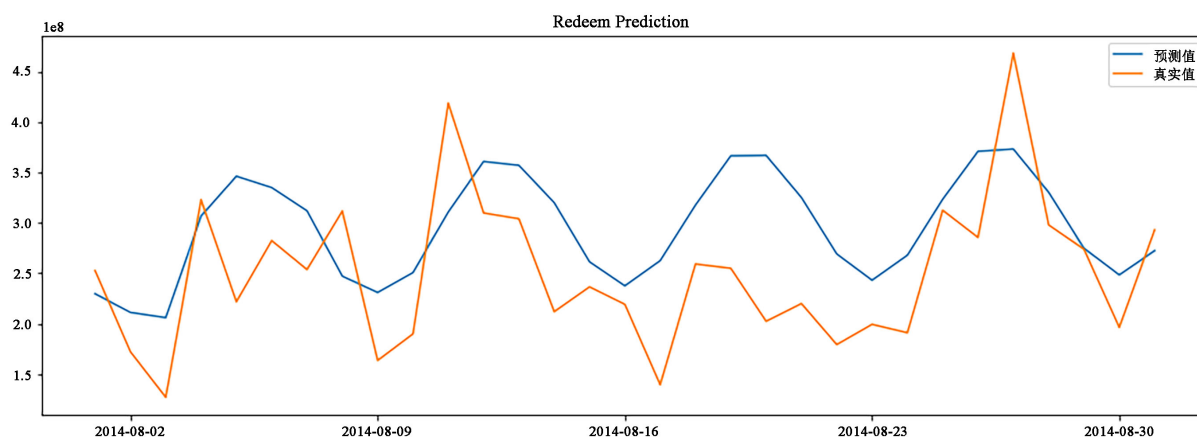


Figure 6. Comparison of redemption model predictions

图 6. 赎回模型预测对比

ARIMA 模型在经过差分处理后提高了精确度, 因为经过差分后数据更加平稳, 规律也更加清晰。由于 ARMA 模型是在统计学的基础上建立起来的, 考虑的特征因素较为单一, 在预测的过程中更为便捷与直观, 同时应用场景广泛, 只要与时间序列相关的数据都可以进行运用, 不需要调整过多的参数就可以迅速给出一个参考数值作为判断依据。但同时其也存在着局限性: 首先, ARMA 需要时序数据是稳定的, 而现实中的数据大多情况下都有自己的趋势, 很难做到完全平稳。其次, ARIMA 模型为线性模型, 无法处理非线性的关系。

除此之外, 经过相关资料的查询与分析, 使用 ARIMA 模型进行金融数据的预测, 例如股票预测往往会出现过拟合的情况, 究其原因, 是因为其预测是存在一些惯性的, 但是股市是存在噪音的, 例如突发事件可能造成市场的急转直下, 所以预测是后知后觉所进行的一个变换, 而这些变换会作为已知值传

入到预测模型中, 数据也会随之发生一些变换, 因此从数据的显示中可以看出预测总是存在滞后性, 导致许多表面看起来很完美的预测结果, 可能只适合于过去, 而无法应对未来所发生的一些未知事件[13]。

在使用 Prophet 模型进行预测时, 将训练组数据代入 Prophet 模型可得到许多参数, 包含了预测值的上下限, 趋势的上下限等数值, 具体的预测结果如图 7 与图 8 所示, 展示了预测的上下波动范围, 同时显示了突变点, 这其中涵盖了许多市场突发的变化所带来的噪音, 随着突变点数量的增多, 模型拟合就变得更加的灵活。

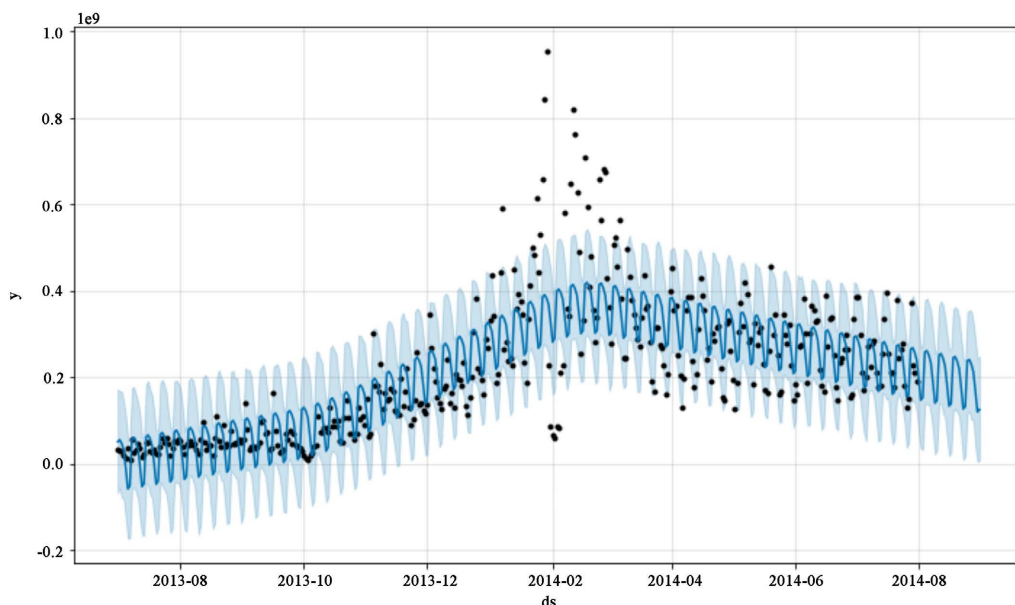


Figure 7. Forecast results of purchase model prophet

图 7. 申购模型 prophet 预测结果

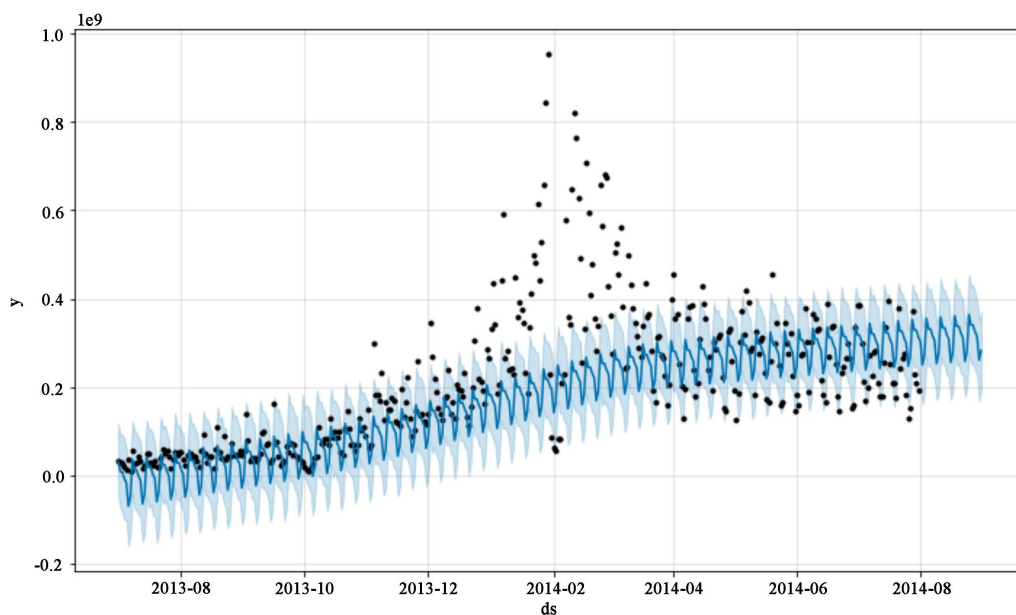


Figure 8. Prediction results of redemption model prophet

图 8. 赎回模型 prophet 预测结果

通过 Prophet 预测结果与 ARIMA 模型对比如图 9 可以看出 Prophet 与 ARIMA 的模型的拟合效果相差不多, 通过特征提取申购量的周期属性(weekday)可以发现用户在一周之内的不同时间点的行为具有一定的规律, 如图 10 所示, 按照 weekday 特征经过统计得出如下图所示的规律, 可以看出周一, 周二的用户申购量最多, 而周五和周六的数量最低。根据余额宝的收益显示规则进行分析, 周五周六所产生的收益用户在下周才能知晓, 因此如果没有特殊情况, 大部分用户会选择在周一或周二进行申购行为而不会选择在周末。

通过分析, 显然 weekday 是一个非常明显的特征, 因此需要在模型拟合的过程中考虑在内。

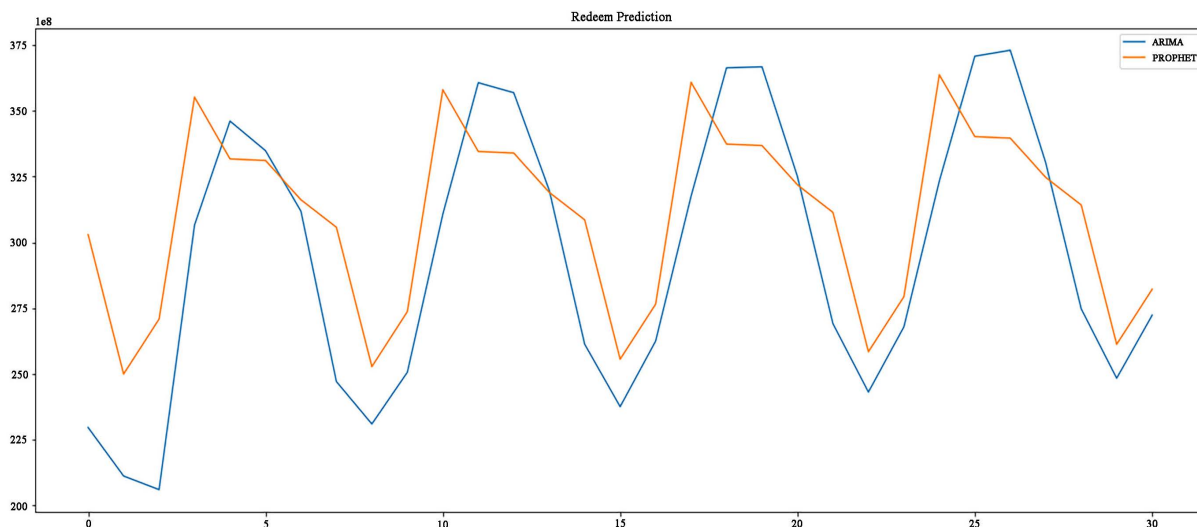


Figure 9. Prediction results of redemption volume by prophet and ARIMA models

图 9. Prophet 与 ARIMA 模型对赎回量预测结果

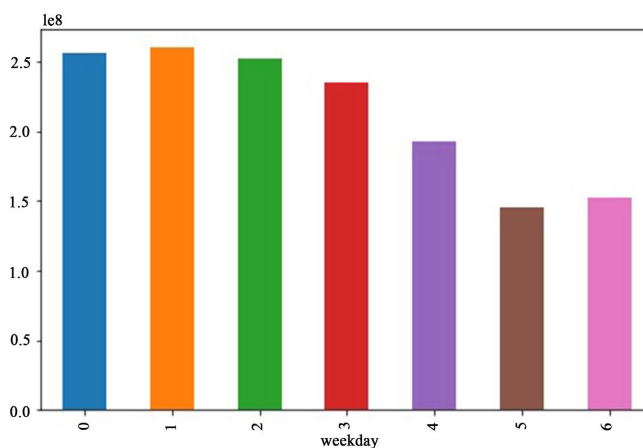


Figure 10. Subscription quantity of users at different times of the week (0 for Monday, 6 for Sunday)

图 10. 用户在一周不同时间的申购数量(0 代表周一, 6 代表周日)

基于上述分析, 由于数据集存在一个非常明显的用户周期性行为, 因此使用了周期因子时间序列模型进行预测。针对资金流入流出的预测, 需要预测的是 8 月份每天的流量, base 可以取 1~30 号的平均流量, 乘上周期因子后得出结果。根据上述计算步骤, 首先需要计算周期因子 weekday, 再计算出 1 号至

30号的均值。需要注意的是,在计算每日均值时会受到周期因子 weekday 的影响,因此需要统计 weekday 在每日出现的频率,剔除周期因子对于每日均值的影响后再将该数值作为 base,最终根据每日的 base 和周期因子进行预测。

由于使用周期因子的方法进行预测需要保证数据集是稳定的,根据时间序列分解可以看出,在2014年4月之后数据集趋于稳定,因此截取14年4月至7月的数据进行周期因子计算,并由此进行相关预测。如图11所示,根据这一段时期内周一至周日的购买与赎回总量与总均值的比值得出周一至周日的周期因子,其中的 purchase_weekday 与 redeem_weekday 分别代表用户申购量与赎回量的周期因子,可以明显比较出周一的数值最大,而周六与周日的数值相对较小。

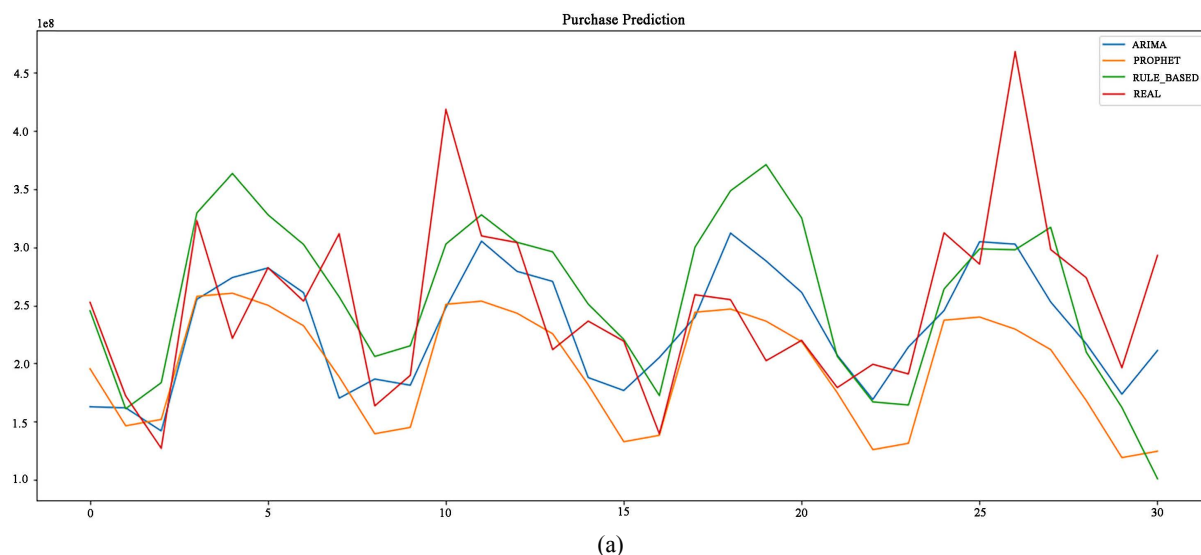
weekday	total_purchase_amt	total_redeem_amt	purchase_weekday	redeem_weekday
0	0	3.051540e+08	1.144828	1.207637
1	1	3.196807e+08	1.199327	1.198612
2	2	3.184404e+08	1.194674	1.132285
3	3	3.133585e+08	1.175609	1.047242
4	4	2.299973e+08	0.862867	0.967552
5	5	1.845509e+08	0.692369	0.681170
6	6	1.857371 e+08	0.696819	0.743258

Figure 11. Calculation results of cycle factor from Monday to Sunday

图 11. 周一至周日的周期因子计算结果

分别统计周一到周日在1号与31号的频次计算出周期因子对于每一天的影响力 day_rate,在按照日期计算每一天的均值 day_mean,根据两者的比值即可得到所需的 base,最后将周期因子乘上 base 可得出最终预测结果。

基于周期因子的时间序列预测模型虽然无法考虑到节假日和突发事件等情况,但基于本文数据集在后没有明显向上或向下的趋势且周期性较强,因此通过计算后发现使用该模型拟合效果良好。将所预测结果与前两种模型进行对比如图12所示。



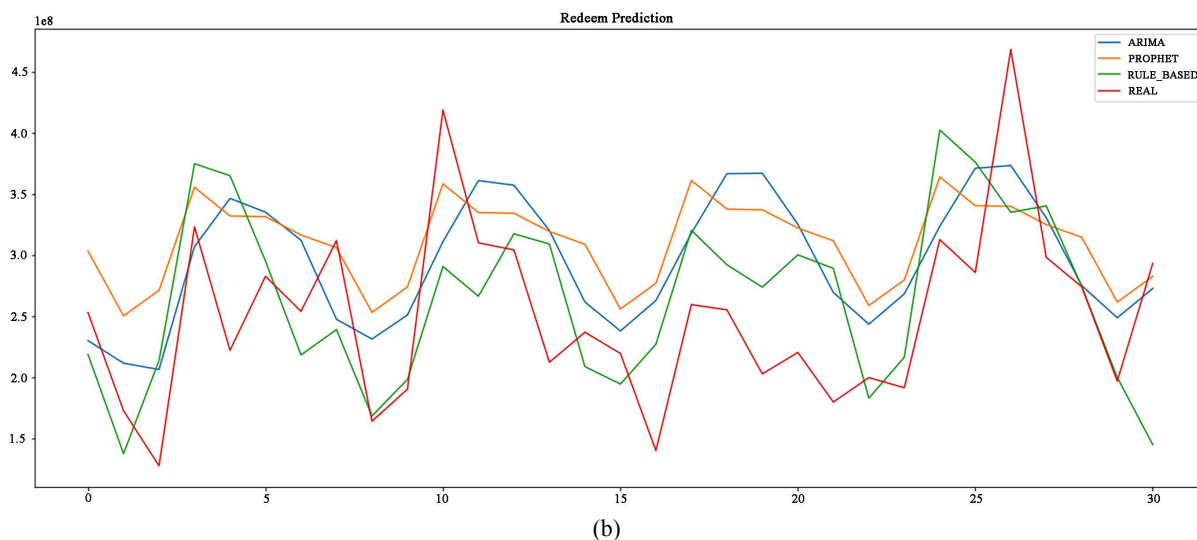


Figure 12. Comparison of model prediction results
图 12. 模型预测结果对比

通过计算 R^2 与均方根误差 RMSE 来检验模型的拟合效果, 如果 R^2 越接近于 1 且 RMSE 越小, 则说明模型的预测效果越好。如表 2 与表 3 所示, 可以看出不论是申购量还是赎回量的预测, 基于周期因子的时间序列模型在进行了数据裁剪, 数据集较小的情况下, RMSE 仍然相对较小, 说明其预测效果是不错的。

Table 2. Comparison of purchase prediction accuracy of three models
表 2. 三种模型申购量预测精度对比

模型	R^2	RMSE
ARIMA	0.25	0.25
Prophet	0.54	0.28
周期因子	0.41	0.22

Table 3. Comparison of redeem prediction accuracy of three models
表 3. 三种模型赎回量预测精度对比

模型	R^2	RMSE
ARIMA	0.31	0.30
Prophet	0.45	0.33
周期因子	0.25	0.29

6. 结论

本文针对余额宝的资金流入流出问题进行了模型拟合与数值分析, 分别使用了 ARMA/ARIMA 时间序列模型、Prophet 模型和基于周期因子的时间序列方法对其进行了统计和拟合, 图 13 和图 14 是对三种模型与真实值的比较, 预测值 1 是 ARIMA 模型的预测结果, 预测值 2 是 Prophet 模型的预测结果, 预测值 3 是周期因子的预测结果。

	日期	真实值	预测值 1	预测值 2	预测值 3
0	20140801	252540858	2.296295e+08	3.029462e+08	2.182880e+08
1	20140802	172250225	2.112063e+08	2.499588e+08	1.370384e+08
2	20140803	127125217	2.060274e+08	2.708788e+08	2.133507e+08
3	20140804	322907524	3.066522e+08	3.552543e+08	3.745535e+08
4	20140805	221706539	3.461415e+08	3.317542e+08	3.647348e+08
5	20140806	282346594	3.349056e+08	3.311775e+08	2.948224e+08
6	20140807	253659514	3.118820e+08	3.162321e+08	2.179700e+08
7	20140808	311648757	2.471051e+08	3.057743e+08	2.387112e+08
8	20140809	163611708	2.309787e+08	2.527870e+08	1.678834e+08
9	20140810	189909225	2.506542e+08	2.737069e+08	1.979324e+08
10	20140811	418603336	3.107064e+08	3.580824e+08	2.903544e+08
11	20140812	309754858	3.607406e+08	3.345824e+08	2.660174e+08
12	20140813	303975517	3.569332e+08	3.340056e+08	3.171520e+08
13	20140814	211939431	3.199044e+08	3.190603e+08	3.087895e+08
14	20140815	236516007	2.613149e+08	3.086025e+08	2.083902e+08
15	20140816	219214339	2.375659e+08	2.556152e+08	1.941034e+08
16	20140817	139564084	2.624959e+08	2.765351 e+08	2.266719e+08
17	20140818	259169016	3.176128e+08	3.609106e+08	3.200753e+08
18	20140819	254929877	3.664001 e+08	3.374106e+08	2.917938e+08
19	20140820	202452782	3.667637e+08	3.368338e+08	2.734537e+08
20	20140821	219963356	3.250149e+08	3.218885e+08	2.999168e+08
21	20140822	179349206	2.691236e+08	3.114307e+08	2.888634e+08
22	20140823	199377531	2.431140e+08	2.584433e+08	1.825151 e+08
23	20140824	191080151	2.679043e+08	2.793633e+08	2.160129e+08
24	20140825	312413411	3.233618e+08	3.637388e+08	4.021412e+08
25	20140826	285478563	3.708084e+08	3.402388e+08	3.758404e+08
26	20140827	468164147	3.730768e+08	3.396620e+08	3.347667e+08
27	20140828	297893861	3.302606e+08	3.247167e+08	3.400234e+08
28	20140829	273756380	2.747714e+08	3.142589e+08	2.740325e+08
29	20140830	196374134	2.483750e+08	2.612715e+08	1.996526e+08
30	20140831	292943033	2.723341 e+08	2.821915e+08	1.445073e+08

Figure 13. Prediction results of three prediction models for purchase volume

图 13. 三种预测模型对于申购量的预测结果

基于上述预测结果主要得出了以下结论:

1) ARIMA 模型、Prophet 模型和基于周期因子的时间序列预测方法对于资金流量数据均有较好的拟合效果, 三种模型的预测精度均在 0.2 左右。但由于其内部逻辑的不同, 基于周期因子的预测方法相较于前两种模型预测的结果相对更加准确, 因为其考虑了序列中非常重要的周期性特征。

2) ARIMA 模型属于探索性预测模型, 是基于统计学基础建立起来的模型, 在预测过程中更为简便, 主要依赖于时间特征进行建模, 很少加入其他参数的影响例如环境限制或周期性变化。因此在进行时间序列预测的前期可以使用该模型迅速得出参考性标准, 但若是要追求模型的精确度, 可能在后期需要添加更多的影响因素以提高模型的泛化性, 防止出现过拟合的情况。

3) Prophet 的优点在于在拟合过程中可以人为设置许多参数, 例如容量值、突变点以及季节性和节假日, 可以预测非线性的周期趋势, 大大增强了模型的泛化性。对于商业领域的金融数据分析, Prophet 可以进行很好的预测, 但可能不适用于一些周期性或趋势性不显著的时间序列。

	日期	真实值	预测值 1	预测值 2	预测值 3
0	20140801	374884735	1.628878e+08	1.954628e+08	2.453163e+08
1	20140802	189092130	1.619184e+08	1.465113e+08	1.611562e+08
2	20140803	173825397	1.420908e+08	1.519927e+08	1.835701 e+08
3	20140804	330640884	2.552046e+08	2.577739e+08	3.294993e+08
4	20140805	394780870	2.739722e+08	2.805084e+08	3.634024e+08
5	20140806	288821016	2.822255e+08	2.501066e+08	3.278816e+08
6	20140807	247646474	2.609463e+08	2.324709e+08	3.023670e+08
7	20140808	233903717	1.702575e+08	1.886268e+08	2.573845e+08
8	20140809	160262764	1.866837e+08	1.396754e+08	2.061097e+08
9	20140810	259534870	1.813714e+08	1.451568e+08	2.152965e+08
10	20140811	331550471	2.481092e+08	2.509380e+08	3.027532e+08
11	20140812	258493673	3.052341 e+08	2.536725e+08	3.278202e+08
12	20140813	261506619	2.792584e+08	2.432707e+08	3.042480e+08
13	20140814	257702660	2.706180e+08	2.256349e+08	2.960202e+08
14	20140815	244551620	1.879540e+08	1.817909e+08	2.511052e+08
15	20140816	215059736	1.768588e+08	1.328394e+08	2.207643e+08
16	20140817	149978271	2.052205e+08	1.383208e+08	1.724655e+08
17	20140818	298499146	2.398047e+08	2.441020e+08	3.000174e+08
18	20140819	266401973	3.121457e+08	2.468365e+08	3.485280e+08
19	20140820	308378692	2.881774e+08	2.364347e+08	3.710191e+08
20	20140821	251763517	2.611745e+08	2.187990e+08	3.251022e+08
21	20140822	246316056	2.072331 e+08	1.749549e+08	2.060963e+08
22	20140823	141412027	1.691544e+08	1.260035e+08	1.670538e+08
23	20140824	130195484	2.142768e+08	1.314849e+08	1.643835e+08
24	20140825	309574223	2.454600e+08	2.372661 e+08	2.640109e+08
25	20140826	306945089	3.047323e+08	2.400006e+08	2.985755e+08
26	20140827	302194801	3.026536e+08	2.295988e+08	2.978451 e+08
27	20140828	245082751	2.528906e+08	2.119630e+08	3.171194e+08
28	20140829	267554713	2.169488e+08	1.681190e+08	2.097128e+08
29	20140830	199708772	1.736732e+08	1.191675e+08	1.622823e+08
30	20140831	275090213	2.111984e+08	1.246489e+08	1.010333e+08

Figure14. The prediction results of three prediction models for redemption volume

图 14. 三种预测模型对于赎回量的预测结果

基于周期因子的时间序列分析在本文的拟合效果最优, 因其加入了一个非常重要的影响因素, 即用户在一周之内不同时间的操作行为具有非常明显的周期性。对于较为平稳的周期性时间序列, 基于周期因子的时间序列方法可以达到比较好的预测结果, 但该方法只适用于一些特定的时间序列, 且只适用于短期预测。

参考文献

- [1] 温玉莲. 基于行业背景差异下的金融时间序列预测方法[D]: [硕士学位论文]. 济南: 山东财经大学, 2021. <https://doi.org/10.27274/d.cnki.gsdjc.2021.000662>
- [2] 郝继升, 任浩然, 井文红. 基于自适应遗传算法优化的 BP 神经网络股票价格预测[J]. 河南科学, 2017, 35(2): 190-195.
- [3] Devi, B.U., Sundar, D. and Alli, P. (2013) An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA

- Model for Nifty Midcap-50. *International Journal of Data Mining & Knowledge Management Process*, **3**, 65-78. <https://doi.org/10.5121/ijdkp.2013.3106>
- [4] Adebisi, A.A., Adewumi, A.O. and Ayo, C.K. (2014) Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, **2014**, Article ID: 614342. <https://doi.org/10.1155/2014/614342>
- [5] 葛娜, 孙连英, 赵平, 万莹. 基于 ARIMA 时间序列模型的销售量预测分析[J]. 北京联合大学学报, 2018, 32(4): 27-33. <https://doi.org/10.16255/j.cnki.ljdxzb.2018.04.006>
- [6] 陶佳, 沈红岩, 高冠东. 基于改进支持向量机的林业资金投资预测方法[J]. 河北农业大学学报, 2012, 35(4): 123-126.
- [7] 赖慧慧. 基于时间序列 Prophet 模型的乘用车消费税预测[J]. 税收经济研究, 2020, 25(1): 34-39. <https://doi.org/10.16340/j.cnki.ssjiyj.2020.01.006>
- [8] 黄莺, 张筠汐. 基于 GM(1,N)-Prophet 组合模型的电商行业销售预测研究[J]. 西南民族大学学报(自然科学版), 2021, 47(3): 317-325.
- [9] Safari, A. and Davallou, M. (2018) Oil Price Forecasting Using a Hybrid Model. *Energy*, **148**, 49-58.
- [10] 曹璨, 黄海, 徐可. 一种基于时间序列算法的资金流入流出预测模型[J]. 微型机与应用, 2017, 36(14): 52-56. <https://doi.org/10.19358/j.issn.1674-7720.2017.14.017>
- [11] 屈妍. 时间序列分析和机器学习在资金流预测中的应用[D]: [硕士学位论文]. 天津: 南开大学, 2021.
- [12] 潘晨. 二种资金流入流出预测模型比较研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2018.
- [13] 李云飞. 金融时间序列方法在股市涨跌序列中的应用及预测研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2020. <https://doi.org/10.27175/d.cnki.gjxcu.2020.001122>