

# 基于 ARIMA 模型的金融品种走势预测技术

王成国, 邓仲元, 陈海文, 蔡志平

(国防科学技术大学 计算机学院, 湖南 长沙 410073)

**摘要:**利用金融品种的历史数据开展数据挖掘,有效预测金融品种的走势,为投资者提供决策导向具有广阔的市场前景和应用价值。文中针对金融品种走势预测的应用需求,深入分析金融品种的时间序列特征,总结出其除了包含常见的非线性、非平稳、动态等特征外,还具有高噪音和非正态等特点。基于求和自回归滑动平均模型,建立金融品种走势预测模型,通过实际数据验证了模型的有效性及其预测的准确性。自回归滑动平均模型可用于金融品种的动态分析和短期预测。

**关键词:**时间序列;差分自回归滑动平均模型;平稳;金融品种推荐;模型预测

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2015)07-0011-04

doi:10.3969/j.issn.1673-629X.2015.07.003

## Financial Variety Trend Prediction Technology Based on ARIMA Model

WANG Cheng-guo, DENG Zhong-yuan, CHEN Hai-wen, CAI Zhi-ping

(College of Computer, National University of Defense Technology,  
Changsha 410073, China)

**Abstract:** The use of data mining based on historical data of financial products can predict the trend of the financial variety, and it has wide application value and market prospects. According to the application demand of the forecast of the financial variety trend, analyze the characteristics of the time series of the financial variety deeply, and summarize that it not only contains the common characteristics of non-linear, non-stationary and dynamic, but also has the characteristics of high noise and non-normal distribution. Based on autoregressive moving average model, establish the forecasting model of the trend of the financial varieties, and verify the validity of the model and accuracy of prediction on the basis of the actual data. The autoregressive moving average model can be used to dynamic analysis and make a short-term prediction of financial products.

**Key words:** time series; auto-regressive integrated moving average model; steady; financial products recommended; prediction of model

## 0 引言

金融投资产品(如黄金)既能给投资者带来大的收益,同时如果投资决策不当也会给投资者带来严重灾难,因此黄金指数走势引起国际社会高度关注。探讨有效的金融产品走势预测方法,为投资者提供决策导向,对黄金走势的预测极具实用价值。

差分自回归滑动平均模型<sup>[1-2]</sup>(Auto-regressive Integrated Moving Average Model, ARIMA)是由英国的 George Box 和 Gwilym Jenkins 在 1970 年初首次提出的基于时间序列的一种著名的预测方法,又被人们称为博克思-詹金斯模型、Box-Jenkins 预测法。这种方法已经在很多领域得到了广泛应用,如工程学领域<sup>[3]</sup>、生物学领域、经济学领域<sup>[4]</sup>和环境卫生学领

域<sup>[5-6]</sup>等。近年来,研究人员开始开展其在金融领域的应用,认为该方法可较好地用于金融品种的预测。

金融品种时间序列是金融运动过程中表现出的依时间为序排列的数据序列,它是可得到的一个真实的有限数据集合。时间序列分析的 ARMA 模型是用来分析这些数据的一种十分有效的方法,其可应用于:

(1) 预测金融品种序列未来的发展趋势;

(2) 分析序列的基本趋势和随机项的组成情况;

(3) 分析特定的样本数据集合,拟和理论模型,尤其是数学模型,进而对模型的整体结构进行详细分析和用实际的数据进行验证探究模型是否符合要求。

文中使用的主要方法是基于时间序列的差分自回归滑动平均模型法,即 ARIMA,来建立金融品种预测

收稿日期:2014-08-22

修回日期:2014-11-26

网络出版时间:2015-06-23

基金项目:国家自然科学基金资助项目(61379145, 61170288)

作者简介:王成国(1987-),男,硕士研究生,CCF 会员,研究方向为数据挖掘;蔡志平,副教授,博士,CCF 高级会员,研究方向为大数据、网络安全。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150623.1028.013.html>

模型。通过对样本的自相关与偏相关函数的观察确定了模型的阶数、自回归系数、滑动平均系数以及偏差等参数。并且分析金融品种变化的规律,进而实现金融品种走势的有效预测,为投资者制定合理的投资决策提供科学依据。通过实验表明,该模型可以用于实际预测且效果良好。

## 1 差分自回归滑动平均模型

常用的预测方法包括趋势外推法、灰色预测法、状态空间模型和差分自回归滑动平均模型法等,每种方法都有其一定的适应范围。趋势外推的基本假设是未来的趋势是由过去和现在连续发展所产生的。当预测对象随着时间的变化呈现出某种上升或下降趋势,没有明显的季节波动,并且能找到一个适合的函数曲线来反映这种变化趋势时,才可以用趋势外推法进行预测。趋势外推法的理论基础是:决定事物过去发展的因素,在很大程度上也决定该事物未来的发展,即使有变化也是细微的;事物发展过程一般都是逐步变化,而非跳跃式的变化,发现并了解事物的发展规律,依据这种规律做推导,就可以预测出其未来趋势和变化。系统因素不完全明确,因素关系不完全清楚,系统结构不完全知道,系统的作用原理不完全明了,像这种信息没有被全部熟知的系统称为灰色系统。在这个系统理论中,使用很少的或不够准确的表示灰色系统行为特征的对原始数据序列作累加生成处理后得出的,用来表示灰色系统内部事物这种连续变化的过程的模型,称之为灰色模型。当序列出现波动比较大的情况时,灰色模型实施的困难程度会加大;状态空间法对原始数据序列的要求是应具有马尔可夫特性,即无后效性,系统将来的变化趋势与过去无关。

ARIMA 模型法目前被称为是最通用的时间序列预测方法,它的基本思想是把时间序列看作一组随机变量,通过分析其自相关性对将来的值做预测。这种方法的优点在于:

- (1)不需要对时间序列未来的发展变化形式作先验假设;
- (2)可以通过多次对模型进行识别与修改,最终获得最佳的模型;
- (3)可以使用计算机软件来实现,操作简单且方便;不仅具有较强的实用性,而且具有较高精确度。

差分自回归移动平均模型法如下所述。

### 1.1 ARIMA 模型预测法的基本思想与该模型的类型

ARIMA 模型方法<sup>[7]</sup>的基本思想是把预测对象随着时间的推移进而形成的数据序列看作一个随机序列,然后再把这个随机序列用一定的数学模型来近似

地描述出来。这个近似的数学模型一旦被识别出来后就可以用时间序列过去的值和现在的值来预测未来的值。

其模型类型分为 ARIMA(  $p, d, q$  )称为差分自回归移动平均模型、AR(  $p$  )自回归模型和 MA(  $q$  )移动平均模型。自回归模型(AR):反映经济变量的当前值与其过去值的关系;移动平均模型(MA):反映经济变量当前值与当前及过去误差项的关系;两者结合的模型为 ARMA。

基本公式为:

$$\hat{y}_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

其中,  $p, d, q$  依次代表的是时间数列中的自回归项数、时间序列成为平稳时所做的差分次数和移动平均项数;  $y_{t-i}$  为  $t-i$  时的实际值;  $\varphi_i$  为该模型的自回归系数;  $e_{t-i}$  为时间序列模型在  $t-i$  时的偏差;  $\hat{y}_t$  为时间序列在  $t$  时的预测值;  $\theta_i$  为该模型的移动平均系数。

### 1.2 预测步骤

预测步骤一般包括三个阶段<sup>[8]</sup>:

第一阶段:对模型的识别。采用一些方法如自相关分析和偏自相关分析等,对原始时间序列的平稳性、随机性进行分析,把自相关分析和偏相关分析准则作为模型阶数判定依据,并建立预测模型 ARIMA。通过运用对其所建模型阶数改变的方法,从中找出使 AIC 和 BIC 值最小的模型,称满足此条件的模型为最佳模型。即把 AIC 和 BIC 的值作为判别模型好坏的依据。

第二阶段:对已建模型的参数进行估计以及验证所建模型。对预测模型的参数进行估计,检验所建的预测模型是否符合要求的依据是通过所建模型与原始的时间序列间产生的偏差数列  $e_t$ 。把  $e_t$  是否具有随机性作为判别所建模型是否满足要求的依据,若具有随机性,则建立的模型满足要求,可以在预测中运用。否则,出现模型需要进一步改进的提醒,遇到这种情形就应该考虑回到上一阶段对模型进行重建。该研究对  $e_t$  是否具有随机性是通过博克斯-皮尔斯中的  $Q$  统计量来验证的,  $\chi^2_\alpha(m)$  如果是统计量且  $m$  满足最大的延迟数,则所建模型通过检验,合乎要求,可以进行对金融品种的预测。

第三阶段:对建立的合乎要求的模型进行实际应用。通过上面各阶段的建模及验证,最终找到最适合的预测模型,然后应用此模型对未来各时期具体的数值做出最合乎实际的预测。差分自回归移动平均模型常用于短期的预测,且精度相对较高,如 1~3 天的预测,因此该研究只对未来三天的黄金价格进行预测,并用 ARIMA 模型法进行连续一个月的预测来进行准确性验证。

## 2 面向金融品种走势预测的 ARIMA 模型

### 2.1 对原始时间序列特征的分析

用 ARIMA 模型法进行预测必须满足的前提条件是:需要预测的时间序列必须要平稳,也即单个样本值要围绕时间序列的平均值做上下浮动,不允许出现显著的下降或上升的趋势,在时序图中表现为所有样本点在一水平线上下随机的波动。如果出现上升或下降趋势,需要对原始序列进行差分平稳化处理。对原始时间序列作时序图、自相关分析图以及偏相关分析图,通过自相关分析图可以知道,原始时间序列尚不合乎此模型对数据序列平稳性的要求。现在要做的就是如何满足模型对数据序列的要求,即实现数据序列的平稳化。这里采用的方法是对原始时间序列进行差分处理。对不平稳的原始数据数列做一阶差分处理,然后通过差分后时间序列 ACF 图的观察,可以知道在  $k=2$  之后的数据序列的自相关系数  $p$  全部出现在 95% 的可信区间中,并且出现系数渐进向零的趋势,把对自相关图的分析作为判断数据序列是否具有平稳性和随机性的准则,原始不平稳的数据序列经一阶差分后满足模型对前提条件的要求(如图 1 所示)。故在此次对预测模型的研究中,有必要对原始的时间数列进行一阶差分处理,待满足模型对平稳性的要求后再进行建模。

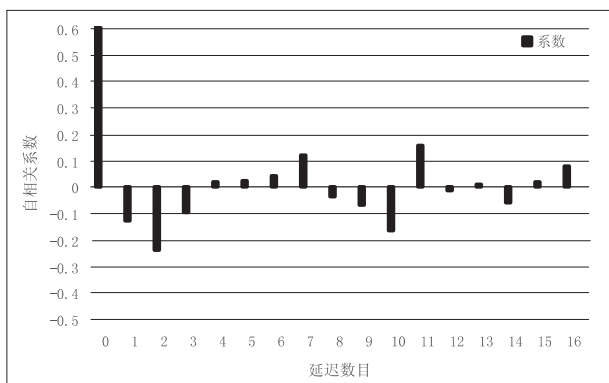


图 1 对原始时间数列一阶差分后的自相关分析图

### 2.2 模型的识别

在无法很精确地确定模型类型的情形时,如果选用  $MA(q)$  或  $AR(p)$  模型在很大程度上会产生与真实模型差别较大的现象。基于这种情况,一般优先考虑使用  $ARIMA(p, d, q)$  模型来拟合原始时间序列,待模型检验和对参数的估计都进行过,然后再确定合适的模型类型。

通常在模型定阶的初期,应该采用阶数由低到高逐次尝试的方式,考虑到如何使阶数的确定过程更加简便,本研究采用吴贤铭和 Pandit 提出的  $ARMA(p, p-1)$  模型法<sup>[9]</sup>。该模型仅有一个变量  $p$ ,采用尝试的方法也会非常简单方便,而且已经确定任何一个平稳

的时间数列,当  $p$  非常大时可以使用模型  $ARMA(p, p-1)$  来对需要的任意精度进行逼近。当然,也可以采用对照分析  $p$  和  $q$  的方式来确定模型的阶数<sup>[10-12]</sup>,对模型的判定可以采用尝试的方法。首先假定一个模型,然后再对其进行分析与检验,推理出其中不合理的成分,然后进行排除、改进等操作。经过分析与检验,最终确定的模型为  $ARIMA(1, 1, 1)$ ,这个模型使全部包含的参数都有意义。

### 2.3 对模型中参数进行估计

依据确定下来的模型  $ARIMA(1, 1, 1)$ ,用 SAS 公司的 JMP 统计软件,按照此操作流程<sup>[13]</sup>:分析→建模→时间序列→ARIMA,进行模型中参数的估计并把模型中的 AIC 和 BIC 计算出来。 $ARIMA(1, 1, 1)$  模型的参数估计的结果  $\varphi_1=0.52$ ,  $\theta_1=0.79$ ,  $e_t=-0.42$ ,经过对所有参数的检验都有意义,其中 AIC 和 BIC 的值分别为 320.121 和 326.353,并且经过模型间(如  $ARIMA(1, 1, 2)$ 、 $ARIMA(2, 1, 2)$ 、 $ARIMA(2, 1, 2)$  等)的 AIC 和 BIC 值进行比较,结果发现模型  $ARIMA(1, 1, 1)$  的 AIC 和 BIC 值是最小的,所选模型又一次得到验证,说明该模型是最符合本次研究的。

最后得出的预测模型为:

$$\hat{y}_t = 0.271y_{t-133} - 0.55e_{t-1} - 0.42$$

### 2.4 对模型进行验证

预测模型建立后,应该对模型的正确性作检验。通常检验是对残差进行的,也就是残差的随机性检验。如果残差具有随机性,那么就认为模型可以用于预测,如果残差不具有随机性,那么就表明模型尚需进一步的修改完善。模型检验通常采用 Box-Pierce 的  $Q$  统计量,统计量  $Q$  服从  $\chi^2_\alpha(K-p-q)$  分布。 $Q$  统计量为:

$$Q = T \sum_{k=1}^K r^2$$

文中把  $Q$  统计量的方法作为模型检验的依据,由  $Q$  统计量的计算方法<sup>[14]</sup>,可知  $Q=8.59$ ,查表得  $\chi^2_\alpha(m)=37.652$ ,  $\alpha=0.05$ ,  $m=16$ 。由于  $Q < \chi^2_\alpha(m)$ ,  $P > 0.05$ ,因此所选模型  $ARIMA(1, 1, 1)$  是满足要求的,可以用于预测。

## 3 实验

### 3.1 数据来源

资料来源于上海黄金行情数据中心,2014 年 5 月 19 日至 2014 年 8 月 12 日的黄金历史价格。由于该数据量大,且相对稳定,因此,文中采用黄金的历史价格来建立时间序列。

每天黄金价格涨跌呈现出明显波动,有很强的非平稳性,每天均出现最高点,但最高点出现的时间不太一致,并没有表现出明显的周期性波动和季节性变化,



见图 2。

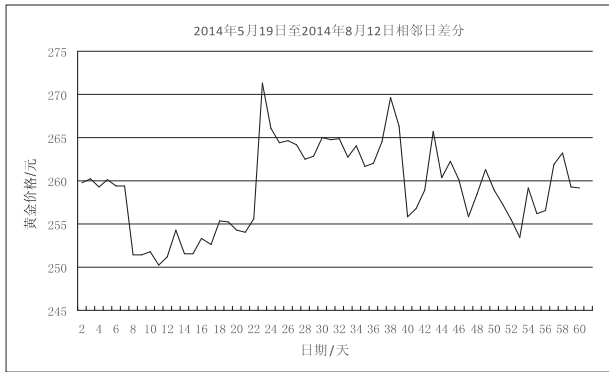


图 2 黄金价格时序图

3.2 利用差分进行原始时间序列的平稳化处理

一般利用  $d$  阶差分来进行趋势明显的时间序列的平稳化,  $d \geq 1$  的整数,如一阶差分。对黄金历史价格的原始时间序列进行一阶差分处理,差分处理后的序列见图 3。

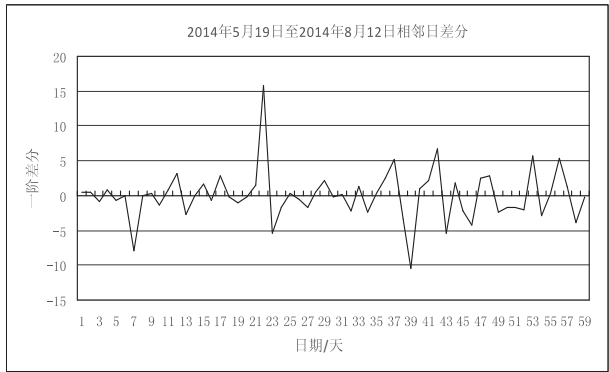


图 3 黄金价格时间序列一阶差分后的序列

由图 3 可以看出,一阶差分后的序列已经平稳。

3.3 预测值与实际值的对比

依据所建立的模型,对 2014 年 5 月 19 日至 2014 年 8 月 12 日黄金价格进行内部代入预测,结果显示预测值与真实值平均相差 3.55,对预测值与真实值作线图后,发现预测数据与实际数据较为吻合,见图 4。图中实线代表实际值,虚线代表预测值。

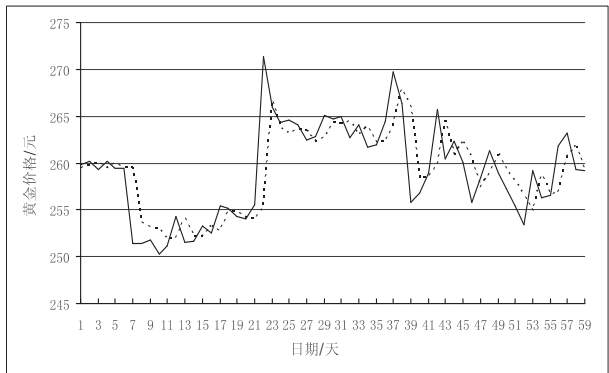


图 4 ARIMA(1,1,1)预测模型拟合图

通过计算实际值和预测值之间的相对误差,得到其相对误差折线图(见图 5)。由图 5 可以观察到大多

数预测数据与实际数据吻合(误差在 0.02% 内),并且通过计算平均误差率,也可以知道,整体的平均误差率,也达到较小的水平(平均误差率为 0.027%)。由此说明,所选模型可以适用于实际数据的预测。从中也能观察到个别预测数据与实际数据误差相差稍大,这种情况一般出现在黄金价格急剧变化的时候。通过加入价格变化斜率等参数,可以减小预测误差。

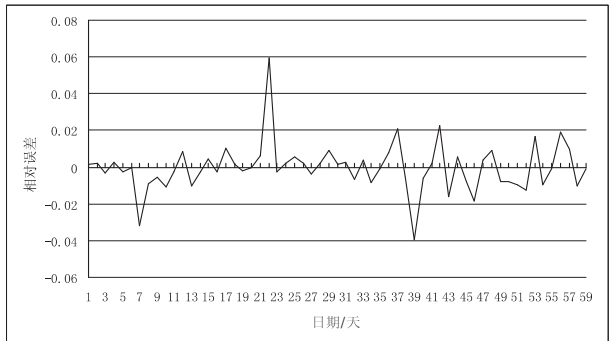


图 5 黄金价格每天的预测相对误差

具体而言,以对 2014 年 8 月 13 日至 15 日的黄金价格进行短期预测为例,预测值与实际值的对比如表 1 所示。其绝对误差控制在 1.5 元以内,相对误差在 0.6% 以下。说明利用时间序列分析工具,可以有效对黄金等金融品种的未来价格和走势进行预测。

表 1 2014 年 8 月 13 日至 15 日黄金  
预测值与实际值的比较

| 时间              | 预测值/元  | 实际值/元  | 绝对误差 | 相对误差/% |
|-----------------|--------|--------|------|--------|
| 2014 年 8 月 13 日 | 259.22 | 260.45 | 1.23 | -0.47  |
| 2014 年 8 月 14 日 | 260.26 | 259.73 | 0.53 | 0.20   |
| 2014 年 8 月 15 日 | 259.29 | 257.88 | 1.41 | 0.55   |

由表 1 可知,通过 2014 年 8 月 13 日至 15 日的数据验证,预测相对误差均小于 2%,预测效果良好。

4 结束语

金融品种预测的影响因素很多,该研究对原始时间序列进行分析,并没有发现其具有明显的周期性,故只有对金融品种进行较长时间连续的动态观察,方可及时准确地了解并掌握金融品种走势变化。文中研究在不考虑周期性因素的影响下,对其建立数学模型并且获得了较为满意的 ARIMA(1,1,1)预测模型,把  $Q$  统计量法作为对所建模型进行检验的依据,结果表明模型在短期预测中效果相当好。尽可能多的获取数据并进行内部代入预测,进一步对模型预测的精确度进行验证。

验证了 ARIMA 模型法在金融品种中的预测效果显著。但是,应用 ARIMA 模型还需注意以下几点:

(1) 不能把一次分析所建立的 ARIMA 模型,看作

(下转第 19 页)

其采样频率普遍较低,且存在诸多无效数据。文中根据出租车运营特性,提出了在定位误差、低频采样、数据量少等因素影响下的路段行程速度的估计方法,并通过实例证明该方法能得出与实际值较为接近的估计值,切实可行。文中所提方法充分利用了现有的数据资源,具有实际意义。但由于出租车行驶情况特殊,计算结果有所偏离,还需要做进一步研究和改进。

参考文献:

[1] Bishop R. Floating car data projects worldwide;a selective review[C]//Proceeding of ITS America annual meeting. [ s. l. ]:[ s. n. ],2004:192-197.

[2] Emam E B, Ai-Deek H. Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability[J]. Transportation Research Record:Journal of the Transportation Research Board,2006,1959(1):140-150.

[3] Quiroga C A. An integrated GPS-GIS methodology for performing travel time studies[R]. [ s. l. ]:[ s. n. ],2002.

[4] 陈 青. 基于GPS浮动车的城市道路交通状态判别技术研究[D]. 西安:长安大学,2009.

[5] Zhu Yanmin,Li Zhi,Zhu Hongzi,et al. A compressive sensing approach to urban traffic estimation with probe vehicles[J]. IEEE Transactions on Mobile Computing,2013,12(11):2289-2302.

[6] 张和生,张 毅,胡东成. 路段平均行程时间估计方法[J].

交通运输工程学报,2008,8(1):89-96.

[7] 姜桂艳,常安德,张 玮,等. 基于GPS浮动车的自然路段行程时间估计方法[J]. 公路,2009(11):87-90.

[8] 朱 鲤,杨东援. 基于低采样频率浮动车的行程车速信息实时采集技术[J]. 交通运输系统工程与信息,2008,8(4):42-48.

[9] Thiagarajan A,Ravindranath L,LaCurtis K,et al. VTrack:accurate,energy-aware road traffic delay estimation using mobile phones[C]//Proceedings of the 7th ACM conference on embedded networked sensor systems. Berkeley:ACM,2009:85-98.

[10] Caveney D S. Method and system for predicting a future position of a vehicle using numerical integration; U. S. ,7865299(B2)[P]. 2011-01-04.

[11] 李筱菁,孟庆春,魏振钢,等. GPS技术在城市交通状况实时检测技术中的应用[J]. 青岛海洋大学学报:自然科学版,2002,32(3):475-481.

[12] 董红召,吴方国. 基于FCM的路段平均行程时间估计[J]. 科技通报,2011,27(3):426-430.

[13] Quiroga C A,Bullock D. Travel time studies with global positioning and geographic information systems; an integrated methodology[J]. Transportation Research Part C Emerging Technologies,1998,6(1):101-127.

[14] 于德新,高学英,杨兆升. 基于GPS数据及车辆运行特性分析的单车路段行程时间估计[J]. 吉林大学学报:工学版,2010,40(4):965-970.

(上接第14页)

始终不变的,只能用于短期预测;

(2)通常对历史统计数据(即原始数据序列)的要求在50个以上;

(3)实际工作中,应获取更多的数据,对已建立的预测模型用新的真实值进行检验,并且要不断向序列中加入新的真实值,来修正或重新拟合更能反映真实情况的黄金走势预测模型。

参考文献:

[1] Box G,Jonkins G. Time series analysis forecasting and control[M]. San Francisco:Holden-Day,1970.

[2] Helfenstein U. Box-Jenkins modelling in medical research[J]. Stat Methods Med Res,1996,5(1):3-22.

[3] 杨叔子. 时间序列分析的工程应用(上下册)[M]. 第2版. 武汉:华中科技大学出版社,2007.

[4] Clements M P,Hendry D F,陆懋祖. 预测经济时间序列[M]. 北京:北京大学出版社,2008.

[5] 时照华,苏 虹,秦风云,等. ARIMA模型在常见呼吸道传染病疫情预测中的应用[J]. 安徽医科大学学报,2013,48(7):783-786.

[6] Cardinal M,Roy R,Lamber J. On the application of integer

valued time series models for the analysis of disease incidence[J]. Statistics in Medicine,1999,18(15):2025-2039.

[7] 徐国祥. 统计预测和决策[M]. 上海:上海财经大学出版社,1998:150-177.

[8] VO THI THANH VAN. 时间序列分析技术的研究[D]. 长沙:湖南大学,2013.

[9] 王振龙. 时间序列分析[M]. 北京:中国统计出版社,1999.

[10] 王 锐. 多种变量选择方法在ARMA阶数确定中的比较[D]. 济南:山东大学,2013.

[11] Sun Zhangzhen,Xu Tianhe,Mo Yijun,et al. Prediction of polar motion based on combination of weighted least-squares and autoregressive moving average[C]//第五届中国卫星导航学术年会论文集-S8 卫星导航模型与方法. 出版地不详:出版者不详,2014.

[12] Chen Jing,He Guozhu,Yu Ping. Prediction of retail price index of Sichuan Province based on ARIMA model[D]. Ya'an:Sichuan Agriculture University,2014.

[13] 罗伯特·卡弗. 探索性数据分析—基于JMP软件[M]. 上海财经大学统计与管理学院,译. 引进版. 上海:上海财经大学出版社有限公司,2013.

[14] Tsay R S. 金融时间序列分析[M]. 第3版. 北京:人民邮电出版社,2014.

基于ARIMA模型的金融品种走势预测技术


作者：

[王成国](#)，[邓仲元](#)，[陈海文](#)，[蔡志平](#)，[WANG Cheng-guo](#)，[DENG Zhong-yuan](#)，[CHEN Hai-wen](#)，[CAI Zhi-ping](#)

作者单位：

[国防科学技术大学 计算机学院, 湖南 长沙, 410073](#)

刊名：

[计算机技术与发展](#) 

英文刊名：

[Computer Technology and Development](#)

年，卷(期)：

2015(7)

引用本文格式：[王成国](#). [邓仲元](#). [陈海文](#). [蔡志平](#). [WANG Cheng-guo](#). [DENG Zhong-yuan](#). [CHEN Hai-wen](#). [CAI Zhi-ping](#)

[基于ARIMA模型的金融品种走势预测技术](#)[期刊论文]-[计算机技术与发展](#) 2015(7)