

基于 ARIMA 模型的股票行情预测

李秀琴, 梁满发

摘要:基于 ARIMA 金融时间序列理论对样本数据进行了 ARIMA 模型识别、参数估计和模型检验,建立了 ARIMA 预测模型,并用建立的模型进行了短期预测和误差分析。模型检验结果显著,预测精度理想。由此论证了用时间序列 ARIMA 模型来预测股票行情是可行的,亦为证券投资分析提供了一个重要的工具。

关键词:股票价格预测;时间序列分析;ARIMA 模型;SAS 建模

中图分类号:F276.3

文献标识码:B

文章编号:1671-6531 (2013) 14-0047-01

股市行情预测方法研究是投资人或金融投资研究者的难题,时间序列模型是一种重要的现代统计分析方法,ARIMA 模型就是一种重要的时间序列模型^[1]。虽然股市行情数据貌似杂乱无章,但大量文献实证研究表明,常常在某一时期市场行情模式也会反复重现,这正是 ARIMA 模型的应用前提条件,因此探索 ARIMA 模型在股票投资预测方面应用是可行的,有价值的。

时间序列分析是决策、预测的主要方法,SAS 软件提供了强大的时间序列分析功能,即 ARIMA()过程^[2]。模型表达形式为 ARIMA(p,d,q),其中 p 为自回归项数,q 为移动平均项数,d 表示差分的阶数。若时间序列是平稳的可直接运用 ARIMA 模型,若时间序列是不平稳的则需要经过 d 阶差分,将非平稳的时间序列转化为平稳的时间序列^[3]。ARIMA 模型数学表达式如下: $\phi(B) \cdot \Delta^d p_t = \theta(B) \cdot \varepsilon_t$ 。式中 $\phi(B)$ 是自回归算子, $\Delta = 1 - B$ 是差分算子, $\theta(B)$ 为移动平均算子^[4]。

本文主要通过应用 ARIMA 时间序列模型对上证指数进行模型识别、模型拟合及检验,并运用拟合模型预测上证指数短期的走势,对预测误差分析检验,判断模型的可靠性及预测效果。

一、实证和预测

(一)数据

1. 样本数据 本文数据来源于 Wind 资讯金融终端,选取上证指数 2005 年 4 月 1 日至 2006 年 3 月 31 日一年间的日收盘指数作为预测模型的建模输入数据。本文将用 ARIMA 模型预测 2006 年 4 月份的收盘指数。

2. 数据处理 本文以日收盘指数作为预测模型的输入数据,即时间间隔的单位按日计算,这样会出现缺值。本文采用了较普遍的线性插值法,若某一天缺值,则以前一天和后一天的收盘指数相加,再除以 2,得出那天的值。连续多天缺值也按这种方法插值。经处理后的时间序列共有 261 个数据,是从 2005 年 4 月 1 日至 2006 年 3 月 31 日,部分样本如表 1 所示。

表 1 时间序列部分样本的数据

日期	收盘指数	日期	收盘指数	日期	收盘指数
2005-4-4	1202.97	2005-4-8	1248.52	2005-4-14	1234.33
2005-4-5	1191.74	2005-4-11	1240.97	2005-4-15	1216.96
2005-4-6	1214.87	2005-4-12	1219.51	2005-4-18	1197.73
2005-4-7	1225.49	2005-4-13	1248.2	2005-4-19	1199.9

(二)建模过程

1. 数据导入 运行 Enterprise Guide 3.0,打开储存在 Excel 中的数据,并转换成 SAS 的数据格式。原来的数据的日期表示形式为:年 年 月 月 日 日 (YYMMDD)。但在转换时,SAS 软件默认的读取输入格式为:MMDDYY。因此要更改日期显示方法,生成 SAS 数据格式的过程为:日线(2005 年 4 月-2006 年

3 月)(sheet1\$)→导入数据→SASUSER.1,运行后生成 SAS 格式的数据文件。

2. 建模过程 利用上述生成的 SAS 数据文件作为输入数据,创建时间序列分析及预测模型——ARIMA 模型。SAS 的“ARIMA 建模和预测”任务分为三个阶段:识别阶段、估计阶段和预测阶段。在此之前先要对该模型设置“任务角色”,把“收盘指数”设为时间序列变量,把“日期”设为时间 ID 变量。

(1)识别阶段 在此阶段首先要设置的是“差分滞后”,这是模型中较为重要的参数之一。若时间序列是平稳的,那么就不需要进行差分,但本文所研究的时间序列为非平稳的,因此对时间序列要进行两次一步滞后的差分,即在差分滞后中填上 (1,1),这时序列的自相关和偏相关函数都呈缓慢的收敛,时间序列经过差分后,近似平稳,数据的个数由 261 个变为 259 个。

下一个要设置的就是平稳性检验及图形和结果。图形和结果主要设置自相关图形的滞后个数,以及图形显示和储存问题。

(2)参数估计阶段 选取了“执行估计步骤”后,就要设置模型定义,也是就该模型的核心内容,需要设置“自回归 p(AR 模型因子)”以及另一参数“移动平均 Vq(MA 模型因子)”。这里和的选取比较复杂,除了会互相影响,率外,根据文献^[5]亦受 AIC 准则要求限制,AIC 的值越小越好。例如选了 1 至 7 作为 q 值,MA1,1、MA1,2、MA1,3 和 MA1,4 参数的 t Value 分别为 -0.01、0.91、-0.10 和 -1.07,由于 t 率太小,所以该项 q 的假设检验并不显著,故可以丢弃这几项。

在不断调试后,得出了一个比较合理的 p 和 q 的值,p=24,26;q=9,10,19,22,30。下一步就是模型选项,本文使用的是“无条件最小二乘法”。

(3)预测阶段 首先选取了“执行预测步骤”,在“观测间的时间”选项中选择“每日,不计周末”,以及确定“要预测的时间间隔”为“20”,即要预测 06 年 4 月份的收盘指数。“置信水平”定为 95%,最后在图形和结果介面中勾取“预测数据”以及“实际值和向前一步预测值”。至此,模型参数设置完成。

(三)预测结果

建模完成的过程图为:日线(2005 年 4 月-2006 年 3 月)(sheet1\$)→导入数据→SASUSER.1→ARIMA

建模和预测→ $\left\{ \begin{array}{l} \rightarrow \text{SASUSER.} \\ \rightarrow \text{HTML-ARIMA 建模和预测} \end{array} \right.$

运行结果显示,AIC 的值为 2254.895,SBS 的值为 2283.35,两者的值也不算太大,说明模型拟合可以接受。通过 ARIMA 建模和预测得到的收盘指数与模型预测的将来值如图 1 所示。

(下转 49 页)

李秀琴/中山火炬职业技术学院公共课部讲师,硕士(广东中山 510640)。

梁满发/华南理工大学理学院副教授,硕士(广东广州 528437)。

来做。橱柜公司根据厨房空间、使用者的爱好习惯,然后进行整体设计,使洗菜、烧饭、备菜等重要区域的空间功能更加合理,降低了制作成本,效果更加整体、美观。卫生间的室内空间分为清洗、浴室、厕所三种功能区域。高档住宅一般是整体浴房,一般的家庭会设置浴帘或隔断等将洗浴与如厕部分分开。厨房与卫生间的顶部都有上下水管道,因此采取铝扣板吊顶处理。吊顶的高度以能遮住裸露的管道为准,一般在2.5米左右。厨房与卫生间的空间立体利用较多。厨房墙面的吊柜、卫生间的水龙头等都使立体空间得到充分的利用。为防止水流到其它空间里,厨房与卫生间的地面都会比其它的空间低2厘米左右。地面铺贴地面砖,门口要安装过门石,消除地面的落差。墙面也要铺贴墙面砖,方便打扫卫生。如是乳胶漆墙面,使用一段时间后,乳胶漆会受潮受污,发生霉变及脱落,也不容易打扫。

阳台是半封闭的空间,阳台的面积在整个家居空间中占据的比例很小,但阳台的作用不可或缺。不仅通风、采光,更是居住者晾晒衣服的最佳场所。理想的户型设计中,客厅与主卧室各有一个阳台。客厅的阳台不仅显得客厅更加明亮、宽敞、开阔视野,而且阳台上可以放置绿萝、吊兰等绿植花卉或水景,打造户内的田园意境,使居住者的心境贴近自然,别有一番情趣。但阳台的盆栽不宜过多,否则会遮挡客厅光线。在卧室的阳台晾晒衣物,可以保证私密性,不会被来访者看见,即使晾晒的衣服杂乱无章,也不会影响整体美观。

在家居室内空间中还有很多零散的剩余空间,这些空间是家居设计中的难点,也是考验设计师设计能力高低的地方。在设计中利用得当,不仅能充分发挥空间的使用率,还能丰富

设计效果,起到重要的点缀作用。例如对于凹凸空间,可设计成小的隔板,摆设工艺品及图书。对于弧形、三角形等不规则的空间,可以设计成搁物架摆放绿色植物,既能充分利用空间,又能净化室内空气,或者设计与其相吻合的家具,充分利用每一空间,丰富空间层次,增加设计效果。

建筑设计是室内装饰设计的基础。在设计室内空间时,必须保护建筑主体结构的完整性。无论是墙面、地面或顶部,承重柱、承重墙及过梁等每一建筑构造都具有重要的承重功能。装饰设计师应熟悉这些基本常识,设计时要仔细分清这些重要的部位,不能单纯追求艺术效果和空间的利用率,而随意砸墙、挖洞及剔槽等,破坏建筑的主体结构,也不能损害墙体预埋的电线、水暖管道、消防管道及燃气管道等。同时室内家居的空间设计在注重整体风格与设计效果的基础上,还要兼顾施工的可行性,力求施工方便,易于操作,不能将设计与施工分离。只有充分考虑到以上方面,巧妙地划分设计每一室内空间,综合运用设计造型、色彩、设计风格及装修的材质等,才能达到最佳的现代室内家居装饰艺术效果。

参考文献:

- [1] 张绮曼,郑曙暘.室内设计资料集[M].北京:中国建筑工业出版社,2005.
- [2] 彭一刚.建筑空间组合论[M].北京:中国建筑工业出版社,2008.
- [3] 韩国建筑世界出版社.居住空间/室内设计方案精选[M].北京:北京科学技术出版社,2010.

责任编辑:何岩

(上接47页)

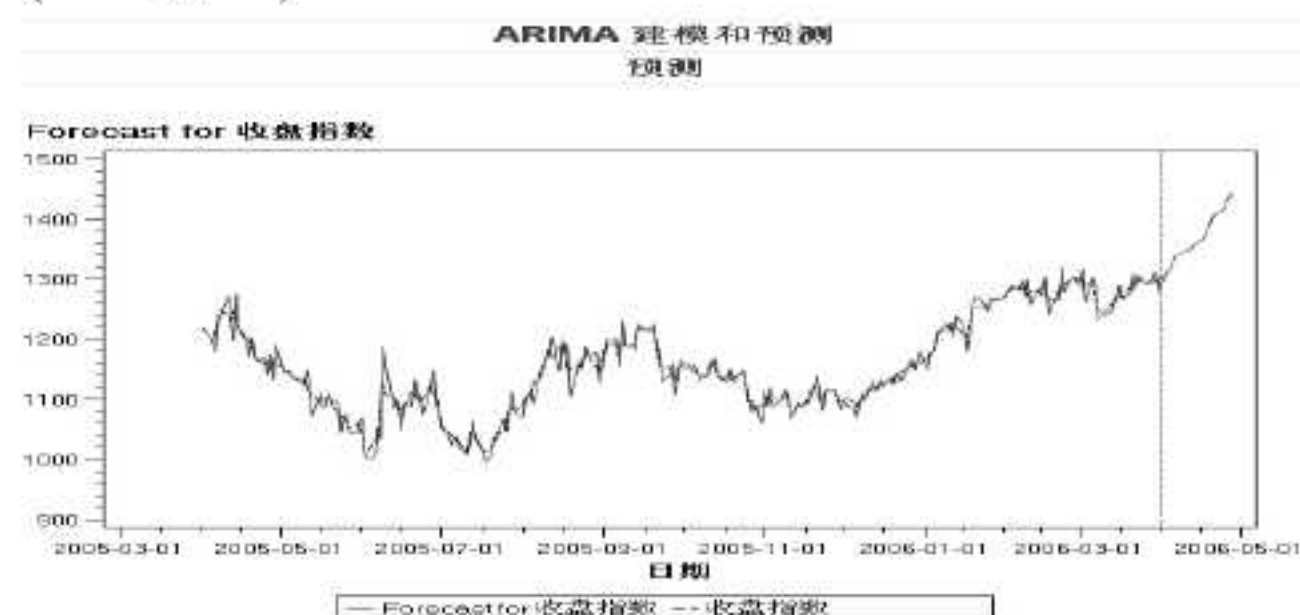


图1 收盘指数及模型预测的将来值

二 模型的预测能力评估

模型预测了06年4月3日至4月28日的上证指数收盘指数。使用Enterprise的图形工具,现将它与实际的收盘指数作比较。通过如下过程:实际指数与预测指数→导入数据→SASUSER.1→线图→HTML线图,得到如图2所示。由此反映出模型的实际预测能力,以及预测模型在股票分析中的存在价值。

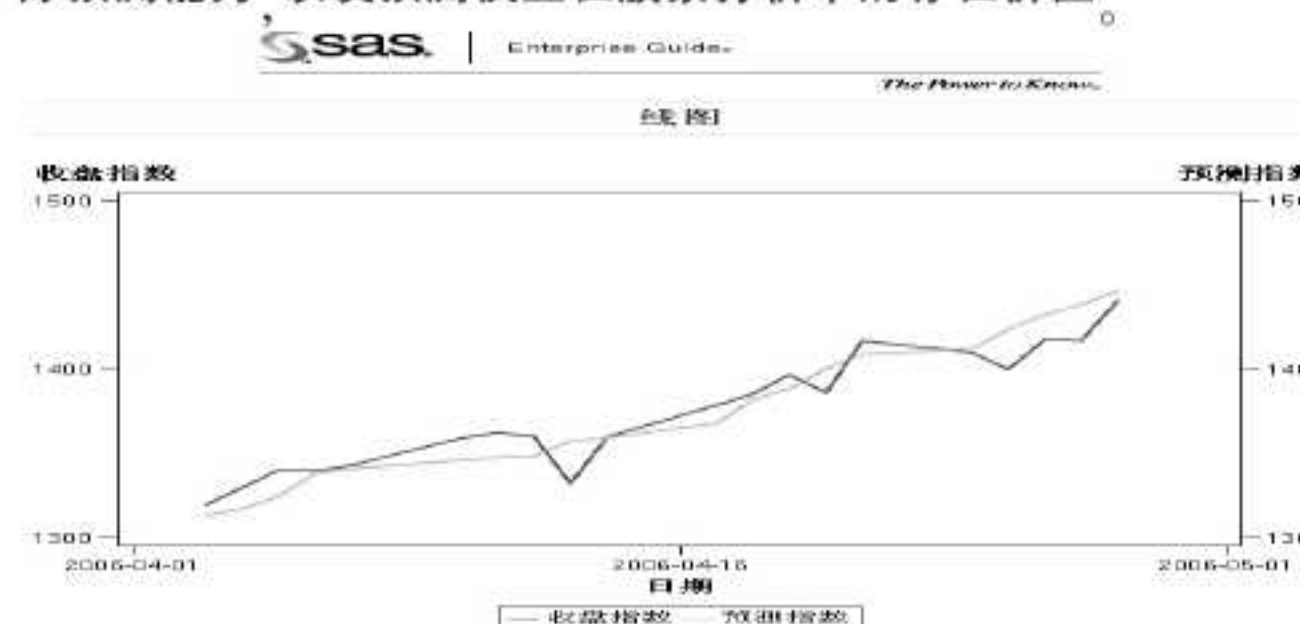


图2 上证指数收盘指数的预测指数与真实指数

从图2中能看出,预测指数虽然与真实指数在一定的差别,但已经能够很好地预测出上证指数收盘指数的基本走势,当中更有部分数值几乎与真实价相同了。

以下将参考文献[5]方法,列表计算对预测与实际值之间的误差、相对误差,如表2所示。

其中:误差=收盘指数-预测指数;相对误差=误差/收盘指数*100%

从表2中看出,相对误差全部都少于2%,拟合程度非常

高,进一步确认了ARIMA模型在短期预测中的准确性。

表2 误差分析

日期	收盘指数	预测指数	误差	相对误差(%)
2006-4-3	1319.47	1313.13467	6.3353313	0.480142125
2006-4-4	1329.8	1317.38893	12.411071	0.933303564
2006-4-5	1340.16	1324.29082	15.86918	1.184125776
2006-4-6	1339.74	1338.00727	1.7327262	0.12933302
2006-4-7	1342.96	1340.94098	2.0190152	0.15034068
2006-4-10	1359.08	1345.95068	13.129315	0.966044331
2006-4-11	1362.23	1347.5659	14.664097	1.076477335
2006-4-12	1360.13	1348.14354	11.986461	0.881273158
2006-4-13	1332.33	1357.42637	-25.09637	-1.883645058
2006-4-14	1359.54	1359.15869	0.3813136	0.028047251
2006-4-17	1378.61	1367.75225	10.857748	0.787586605
2006-4-18	1385.11	1381.8041	3.3059037	0.238674452
2006-4-19	1396.7	1388.30175	8.3982494	0.601292289

三 结论

本文选取上证收盘指数作为研究对象,使用SAS软件操作了ARIMA模型建立过程,并借此来探寻股市的预测方法。本文通过ARIMA模型各种搭配反复试算,建立精度较理想的预测模型,提供了能进行股票指数短期预测的量化投资方法。由于我们选取的数据不够充分,实证结果还存在局限性,结果仅作参考。

基金项目:国家社科基金项目(编号:11XGL009);教育部社科项目(编号:10YJA630207);广东省中山市科技计划基金资助项目(编号:20114A223)

参考文献:

- [1] 王波,张凤玲.神经网络与时间序列模型在股票预测中的比较[J].武汉理工大学学报(信息与管理工程版),2005,27(6):69-73.
- [2] 贾勇宁.分析、预测方法在决策支持中的应用[J].铁路通信信号,2004,40(5):12-14.
- [3] 厉雨静,程宗毛.时间序列模型在股票价格预测中的应用[J].商场现代化,2011,(33):61-63.
- [4] 赵志峰.对建立中国股票价格指数时间序列模型的探讨[J].统计与信息论坛,2003,18(1):66-69.
- [5] 李民,邹捷中,李俊平,梁建武.用ARMA模型预测深沪股市[J].长沙铁道学院学报,2002,18(1):81-87.

责任编辑:姚旺