



基于 ARIMA 模型的体育彩票销量预测

——以上海为例

吴 殷, 李 海

摘 要: 对未来体彩销量预测分析, 可以为体彩产品设计、营销推广提供科学依据。本文以上海为例, 采用 ARIMA 模型对 2009 年 1 月—2013 年 3 月的上海体育彩票的月度销售数据进行时间序列分析, 并预测出 2013—2015 三年上海体育彩票的销量, 从而对上海体育彩票的销量增长趋势进行分析。

关键词: ARIMA 模型; 体育彩票销量; 预测; 时间序列

中图分类号: G80-05 文献标志码: A 文章编号: 1006-1207(2013)05-0023-04

Sales Volume Prediction of the Sports Lottery Based on ARIMA Model ——Taking Shanghai as an Example

WU Yin, LI Hai

(School of Economy & Management, Shanghai University of Sport, Shanghai 200438, China)

Abstract: Prediction of the sales volume of the sports lottery in future may provide scientific reference for the design of the sports lottery products and marketing & promotion. Taking Shanghai as an example and using the model of ARIMA, the paper makes a time series analysis of the monthly sales data of Shanghai sports lottery from January 2009 to March 2013. It predicts the sales volume of Shanghai sports lottery in the years of 2013-2015 and makes an analysis of the growth trend of the sales volume of Shanghai sports lottery.

Key words: ARIMA model; sales volume of sports lottery; prediction; time series

自 1994 年起在全国统一发行, 体育彩票伴随着我国经济社会的快速发展而日渐成长壮大。中国体育彩票 2012 年总销量超 1 104 亿元。在短短 18 年中, 体彩行业保持了年均 37% 以上的增长率, 累计发行超过 5 600 亿元。体育彩票销量的快速增长为国家和社会筹集了大量的公益金, 增加了地方税收, 提高了国家的财政能力。同时, 体育彩票带动了相关产业发展, 为社会提供了大量的就业岗位。但是发展中伴随着各种问题的存在, 如市场开发和营销工作还处在初级阶段, 何时何地采用何种营销手段、该淘汰哪些产品以及该推广哪些新产品的分析和研究还需要进一步研究。而这些问题解决都建立在对未来销量预测的基础上, 因此做好未来体彩销量预测分析, 为体彩产品设计、营销推广提供科学依据, 对于体育彩票产业的健康发展具有重要的理论与现实意义。

影响体育彩票销量的因素是多方面的, 如经济水平、市场规模、居民人口、居民收入、节假日波动、体育赛事等, 这些因素共同影响了体育彩票的销量。同时, 体育彩票月度销量具有明显整体趋势变动性和季节波动性, 即呈现时间序列二重趋势变化特点。对时间序列进行预测的模型和方法较多, 如移动平均法、指数平滑法、ARIMA 模型。由于 ARIMA 模型在预测过程中考虑了随机波动的干扰和时间序列上的依存关系, 因此 ARIMA 模型在时间序列预测中应用较为广泛。本文拟采用 ARIMA 模型对 2009 年 1 月

—2013 年 3 月的上海体育彩票的月度销售数据进行时间序列分析, 并预测出 2013~2015 三年上海体育彩票的销量, 从而对上海体育彩票的销量增长趋势进行分析。

1 ARIMA 模型的基本原理

ARIMA 模型依据变量本身的变化规律, 利用外推机制描述时间序列的变化, 着重分析经济时间序列本身的概率或随机性质, 而不考虑以经济理论为依据的解释变量的作用, 适用于宏观经济影响指数时间序列的预测。ARIMA 模型全称为自回归移动平均模型, 有 4 种基本类型: 自回归 (AR) 模型、移动平均 (MA) 模型、自回归移动平均 (ARMA) 模型以及差分自回归移动平均模型 (ARIMA)。其中 ARIMA (p, d, q) 称为差分自回归移动平均模型, AR 是自回归, p 为自回归项, MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数。

1. AR(p) 模型: P 阶自回归模型, 满足下面的方程:

$$u_t = c + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \varepsilon_t, t = 1, 2, 3, \cdots, T$$

2. MA(q) 模型: q 阶的移动平均模型, 满足下面的方程:

$$u_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, t = 1, 2, 3, \cdots, T$$

收稿日期: 2013-8-31

作者简介: 吴殷, 男, 副教授。主要研究方向: 体育赛事评估、体育博彩、运动心理研究。

作者单位: 上海体育学院 经济管理学院, 上海 200438



3. ARMA(p,q) 模型将纯 AR(p) 与纯 AR(q) 结合, 得到一个一般的自回归移动平均方程 ARMA(p,q) :

$$u_t = c + \phi_1 u_{t-1} + \dots + \phi_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, t=1,2,3,\dots,T$$

4. ARIMA(p,d,q) 模型: 通过对不平稳的时间序列进行 d 阶差分, 将其转化 d 为平稳时间序列, 然后建立 ARMA(p,q) 模型。设 μ_t 是 d 阶单整时间序列, 即 $\mu_t \sim I(d)$, 则:

$$\omega_t = \Delta^d u_t = (1-L)^d u_t$$

ω_t 为平稳时间序列, 即 $\omega_t \sim I(0)$, 于是可以对 ω_t 建立 ARMA(p,q) 模型:

$$\Phi(L) = C + \phi_1 \omega_{t-1} + \dots + \phi_p \omega_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

上式表明如果一个序列是单整序列, 那么该序列可以由自身的滞后值以及随机扰动项来解释。即如果该序列平稳(它的行为并不会随着时间的推移而变化), 那么就可以通过该序列过去的行为来预测未来, 其大致模型如图 1。

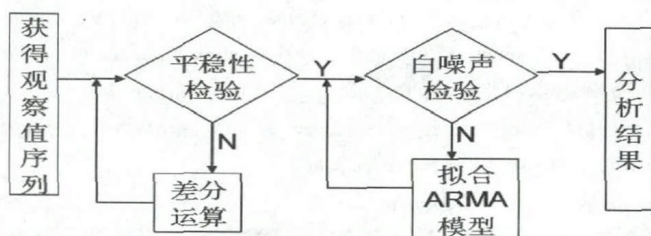


图 1 ARIMA 模型流程图

Figure 1 Flow Process of ARIMA Model

2 上海体育彩票销量时间序列模型的建立及未来 3 年销量预测

本文利用 Eviews7.1 对上海体彩月度销量进行建模, 并对未来三年的销量数据进行预测, 得出将来销量数据的趋势。

2.1 时间序列观察

在建立模型之前, 需要对数据进行分析, 从中发现数据本身的特点和联系。首先, 可将销售数据按时间先后顺序绘制折线图。由图 1 可见体育彩票月销量呈明显的周期性波动, 每年的 1 或 2 月份销量大幅回落, 在 2 月或 3 月销量大幅增加, 这充分显示了春节对体育彩票销量的影响很大。而 7、8 月份通常是体育彩票的销售淡季, 这与国外各大联赛均为休赛期间有关, 也充分显示了赛事对体育彩票销量的较大影响。正是因为这种影响造成的波动较大, 也增加了体育彩票月销量预测的可能误差。其次, 从图 2 可以看出, 虽然每年体育彩票销量波动较大, 但总体看有同期逐渐递增趋势, 这表明体育彩票的月总销量数据呈现出显著的季节性周期规律和明显的总体递增趋势, 具备时间序列二重趋势变化特点, 显示该销量数据序列是为非平稳序列, 需对其进行平稳化处理。

2.2 对序列进行平稳化处理

为了使体育彩票月度销量能更好地符合 ARIMA 模型的



图 2 2009 年 1 月 -2013 年 3 月上海体彩销量趋势图

Figure 2 Sales Volume of Shanghai Sports Lottery in January 2009 to March 2013

条件, 我们先对该体育彩票销量序列 X 进行对数转换, 即 $Y = \ln(X)$ 。经过发现二阶差分后, 通过了单位根检验(图 3), 表明 $\ln(x)$ 在二阶差分后平稳, $d=2$ 。因此可初步确定的模型为 ARIMA(p, 2, q)。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.565497	0.0000
Test critical values:		
1% level	-3.588509	
5% level	-2.929734	
10% level	-2.603064	

图 3 $\ln(X)$ 序列的二阶差分单位根检验

Figure 3 Second Order Differential Unit Root Test of $\ln(X)$ Series

2.3 模型的 p 和 q 的确定

模型的 p 和 q 通常由自回归系数和偏自回归系数来确定。通过观察偏自相关图可以确定 p 为 1 或者 5。而通过观察自相关图(图 4)判断可以确定 q 为 1。由于 $d=2$, 所以最优模型可能是 ARMA(1, 2, 1) 或者 ARMA(5, 2, 1)。

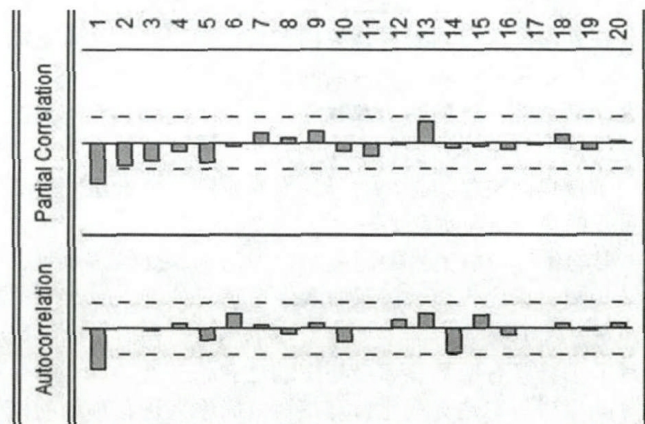


图 4 序列 $Y = \ln(X)$ 的二阶差分后自相关和偏自相关分析图

Figure 4 After Second Order Differential Autocorrelation and Partial Autocorrelation Analysis of the Series $Y = \ln(x)$

2.4 模型的选择

下面对比两组模型。由图 5 的 ARMA(5, 2, 1) 可知其调整后的 R^2 为 0.792 大于图 6 中的 ARMA(1, 2, 1) 模型的

0.589, 而 AIC 和 SC 值分别为 -0.731 和 -0.487, 分别小于图 6 中的 -0.201 和 -0.123, 可以认为 ARMA(5, 2, 1) 更为合适。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.900229	0.141271	-6.372362	0.0000
AR(2)	-0.997852	0.160902	-6.201604	0.0000
AR(3)	-0.895635	0.175392	-5.106460	0.0000
AR(4)	-0.790686	0.174995	-4.518336	0.0001
AR(5)	-0.639358	0.154676	-4.133534	0.0002
MA(1)	-0.955735	0.036930	-25.87931	0.0000
R-squared	0.815858	Mean dependent var		0.016906
Adjusted R-squared	0.791628	S.D. dependent var		0.345383
S.E. of regression	0.157659	Akaike info criterion		-0.730635
Sum squared resid	0.944547	Schwarz criterion		-0.487336
Log likelihood	22.07397	Hannan-Quinn criter.		-0.640408
Durbin-Watson stat	2.366199			

图 5 ARMA(5, 2, 1) 模型参数估计与检验结果

Figure 5 Parameter Estimation and Test Results of the Model ARMA(5, 2, 1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.353880	0.158519	-2.232410	0.0305
MA(1)	-0.944515	0.036236	-26.06593	0.0000
R-squared	0.597450	Mean dependent var		0.010692
Adjusted R-squared	0.588699	S.D. dependent var		0.334350
S.E. of regression	0.214428	Akaike info criterion		-0.200913
Sum squared resid	2.115047	Schwarz criterion		-0.122947
Log likelihood	6.821923	Hannan-Quinn criter.		-0.171450
Durbin-Watson stat	2.028583			

图 6 ARMA(1, 2, 1) 模型参数估计与检验结果

Figure 6 Parameter Estimation and Test Results of the Model ARMA(1, 2, 1)

2.5 模型的建立

根据图 5 和图 6, 我们可以确定 ARMA(5, 2, 1) 为体育彩票销量的最佳预测模型, 该模型的参数估计及模型的相关检验结果如图 5, 其展开式为:

$$\text{ddlnxt} = -0.900229\text{ddlnxt-1} - 0.997852\text{ddlnxt-2} - 0.895635\text{ddlnxt-3} - 0.790686\text{ddlnxt-4} - 0.639358 + \mu_t - 0.955735\mu_{t-1}$$

2.6 模型的诊断检验

对模型作拟合后得到 X 的残差, 对残差做单位根检验 (图 7) 和进行自相关和偏相关分析 (图 8)。图 7 和图 8 表明模型已经提取了有规律的信息, 模型拟合程度较好。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.868019	0.0000
Test critical values:		
1% level	-3.592462	
5% level	-2.931404	
10% level	-2.603944	

图 7 残差的单位根检验结果

Figure 7 Unit Root Test Results of the Residual

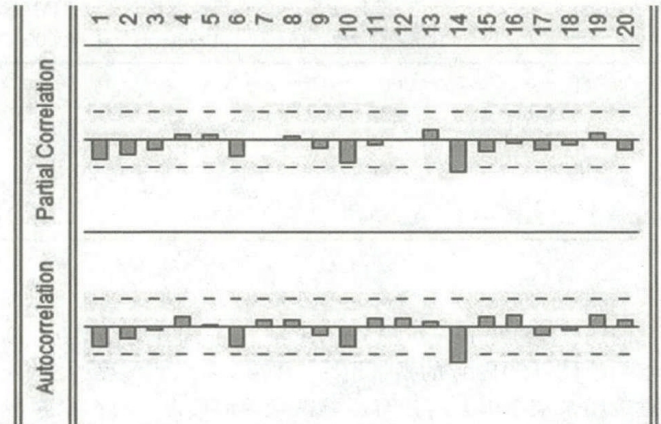


图 8 残差的自相关和偏自相关分析

Figure 8 Autocorrelation and Partial Autocorrelation Analysis of the Residual

2.6 模型的预测

利用 Eview7.1 的预测功能对未来三年的各月总销量进行预测, 结果如图 9 所示。2009-2015 的年度销量的预测结果如表 1, 预计 2013 年上海体育彩票销量为 30.83 亿; 2014 年销量为 35.27 亿, 2015 年销量为 41.38 亿。

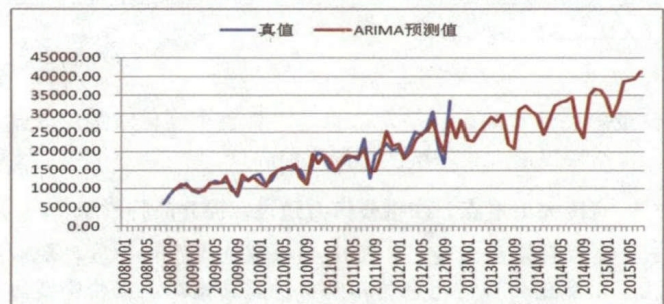


图 9 总销量的 ARIMA 预测值与真值对比

Figure 9 Comparison between the ARIMA Predicted Value and Real Value of the Total Sales Volume

通过模型预测, 可以发现 2009~2012 年中的 4 个预测年度的误差百分比都小于 2%, 另外我们的预测中平均绝对百分误差为 3.71, 表明预测精度较高, 可用于实际预测。

3 结论

通过体育彩票销量的时间序列分析能够揭示潜在的变化趋势, 因此在体育彩票的管理和决策工作中, 销量的时间序列分析是一项重要的内容。目前 ARIMA 模型已广泛应用于人口预测、经济预测等领域, 而应用于体育彩票领域的预测研究还未见报道。

本文通过对 2009 年 1 月 -2013 年 3 月的上海体育彩票的月度销售数据进行时间序列分析, 建立了 ARIMA(5, 2, 1) 预测模型, 预测精度较高, 表明 ARIMA 模型也能较好的应用于体育彩票销量的预测。

预测结果表明预计 2013~2015 年上海体育彩票销量将继续平稳增长, 年均增长 5 亿, 年均增速 16.3%。同样采用 ARIMA 模型对电脑型和即开型彩票的销量进行预测,



表 1 上海体育彩票年度销量的 ARIMA(5,2,1) 模型预测 (2009~2015) (单位: 万元)
Table ARIMA(5,2,1) Predicted Value(2009~2015) of Shanghai Sports Lottery Annual Sales Volume

年份	2009	2010	2011	2012	2013	2014	2015
实际值	119176.1	157584.5	204513.5	263933.2	308326.34	352738.78	413815.86
预测值	119966.1	155126	207290.9	262908.7			
误差百分比	-0.66%	1.58%	-1.34%	0.39%			

其结果表明电脑型彩票在这 3 年的销量也将以年均增长 5 亿的速度继续增长, 年均增速 17.4%; 而即开型彩票在这 3 年的销量与 2012 年的销量相比, 并未出现较大幅度的增长。这表明未来 3 年上海体育彩票销量的增长点主要在电脑型彩票。

参考文献:

[1] 王振龙. 时间序列分析[M]. 北京: 中国统计出版社, 2003

- [2] 何书元. 应用时间序列分析[M]. 北京: 北京大学出版社, 2003
- [3] 杨亚莉, 程林林, 张永韬. 体育彩票销量的计量模型及促销策略研究——以四川省为例[J]. 成都体育学院学报, 2012, 38(9): 1-7.
- [4] 李刚. 彩票人均销量的决定因素和我国彩票市场发展趋势的预测[J]. 体育科学, 2006, 26(12): 38-45.
- [5] 徐国祥统计预测和决策[M]. 上海: 上海财经大学出版社, 1998
- [6] 李子奈, 叶阿忠. 高级计量经济学[M]. 北京: 清华大学出版社, 2000

(责任编辑: 孙胜男)

(上接第22页)

记录彩民消费偏好、基本信息等, 提供个性化服务。以此回报给彩民最大最深层次的价值和满意。

3.5 宣传突出重点, 注重公信力建设, 提升品牌形象

体育彩票在宣传过程中应当不仅注重宣传力度, 更应注重重点宣传和特色宣传相结合的宣传策略。宣传中要体现体彩的差异性, 并逐步打造体彩在彩民中的口碑, 利用口碑效应, 带动品牌形象的建立。宣传中要体现体彩的体育特色和公益因素, 将文化内涵黏附于品牌形象。另外, 宣传中要体现玩法的透明度和公平性, 提升体彩的公信力。让彩民真正了解到体彩作为国家彩票, 目的是支持我国各项公益事业。

3.6 完善培训机制, 提升培训效果

制定完善的培训机制是提高销售人员素质的必经之路。培训应该以销售员为中心, 了解销售员的需求, 并区分不同层次的不同需求, 做到“因材施教”的针对性培训。并

且合理规划培训时间、培训课程, 构建培训评估体系。在培训中融入更多实践内容, 注重培训后的评估和效果测量。

参考文献:

- [1] 菲利普·科特勒. 营销管理[M]. 北京: 中国人民大学出版社, 2005.
- [2] 李海. 新编体育博彩概论[M]. 上海: 复旦大学出版社, 2013.
- [3] 李海. 体育彩票: 公益事业的助推器, 体育事业的生命线[J]. 体育科研, 2012, 33(3): 7.
- [4] 刘鹏. 2010年全国体育彩票工作会议, 为体育彩票实现跨越式发展奠定基础——国家体育总局局长刘鹏. 彩票快讯[Z]. 2010, 2: 7-13.
- [5] 曾雯彬, 孙杨, 孙东辉. 我国体育彩票发展现状及发展趋势研究[J]. 体育科研, 2012, 33(3): 14-17.

(责任编辑: 孙胜男)